Chapter 5

Experiments on Chord Recognition

5.1 Goals of Computer Experiments

In order to investigate the properties of the model from the previous chapters, we have performed a series of computer experiments on chord recognition with synthesized data.

First of all we have implemented the following simple correlation approach.

- **Basic (Simple Correlation) Approach**

  Recall that the recognition of chords in our model is based on recognizing interval relationships between repetitive subspectra. We distinguish between two types of intervals: *harmonic*, or *vertical intervals* between tones of the same chord, and *melodic*, or *horizontal intervals* between tones of different chords. A chord is understood to be an acoustical contour which is drawn by a tone spectral pattern in the frequency domain. Since this contour is constituted by harmonic intervals, the recognition of harmonic intervals is fundamental in the recognition of separate chords. Similarly, a polyphonic part is understood to be a dynamical trajectory which is drawn by a tone spectral pattern in the frequency domain versus time. Since this trajectory is constituted by melodic intervals between every two chords, the recognition of melodic intervals is fundamental in tracking polyphonic voices.

  Harmonic intervals are found by the autocorrelation of the chord spectra with the $\log_2$-scaled frequency axis, and melodic intervals are found by the correlation of spectra of the chords confronted. The correlated groups of partials are interpreted as tone patterns. Therefore, every chord can be recognized either being taken separately, by recognizing harmonic intervals, or being confronted to another chord, by melodic intervals.

The recognition of chords by finding either harmonic or melodic intervals by correlation analysis of chord spectra constitutes the basic element of our model.

A particular difficulty in the recognition of both separate chords and chord progressions is preventing "missed" tones and rejecting "false" tones. Usually, the spectrum of a missed tone is contained in some larger correlated group of partials which is interpreted as another tone. False tones result from accidental correlations caused by structural coincidences in the chord spectra. To a great extent, this difficulty can be overcome by complementing the basic approach by the further improvements.

- **Extensive (Decision Making) Approach**

  Since each chord is recognized in several independent ways, being taken separately and being confronted to its neighbors, the results of these different recognition procedures are processed by a decision making model.

  At first, for every recognition procedure we formulate a hypothesis on the compound of the chord. For this purpose, each correlated group of partials from the intervals found is represented by "conventional pitch" (e.g. by the lowest partial from the group of partials correlated), and the chord is identified with certain "notes". Next, the final decision about accepting or rejecting a note is made with taking into account the frequency of this note in these hypotheses.

  In our experiments we have formulated three hypotheses concerning the compound of each chord, obtained from the analysis of the chord's harmonic intervals, from the analysis of melodic intervals between the chord and its predecessor, and from the analysis of melodic intervals between the chord and its successor. One can consider more hypotheses by confronting the chord to all other chords in the given progression, but even the three hypotheses mentioned are sufficient for providing a considerable improvement in the recognition reliability.

  This approach is extensive, since it is based on multihypotheses decision making by taking into account as much information about the chord in the context as possible.

- **Intensive (Structural) Approach**

  In order to reveal missed tones and recognize accidental correlations in the chord spectrum, the structure of correlated groups of partials is analyzed. Since accidental correlations arise randomly, the structure of accidentally correlated groups of partials is also random, whereas the spectral structure of true tones is stable. This observation is used in order to reject accidentals and recognize missed tones which are masked by larger groups of partials.

For this purpose, different correlated groups of partials are compared in order to recognize repetitive structures in them. The multicorrelation analysis of a chord spectrum described in Chapter 4 provides a means for finding stable groups of partials in a spectrum of a single chord and for rejecting accidental pairwise correlations. Obviously, this approach can be extended to finding repetitive voice spectra in the progression of chords in order to recognize polyphonic parts.

This approach is intensive, since it is based on profound analysis of a given spectral structure and self-organization of data.

Roughly speaking, under the decision making approach we firstly identify two-tone intervals with notes, and then reject some of the notes. On the contrary, under the structural approach we firstly reject some of the intervals, and then identify the remaining ones. Therefore, under the former approach the rejection criterion operates on notes and is based on accounting their frequency, whereas under the latter it operates on the structure of spectral patterns correlated and is based on the simplicity principle.

According to the above classification, we have performed two series of experiments on chord recognition. In the first series based on the recognition of two-tone intervals we have considered chord sequences. For this series of experiments we formulate several important conclusions about the performance of the basic model of interval recognition.

The second series of experiments is based on finding multicorrelated subspectra in spectra of separate chords. For this series of experiments we formulate conclusions about the correspondence of recognition results to certain properties of human perception.

In Section 5.2, "Example of Chord Recognition", we explain the operation of the model with a simple example. We trace the correlation analysis of Boolean spectra of two two-tone chords, according to the model from the previous chapter. We show how accidental patterns can arise and how they can be efficiently recognized in the test experiments. Then we introduce several characteristics with which the performance of the model is studied in the sequel.

In Section 5.3, "Testing the Simple Correlation Approach", we describe the procedures used in computer experiments on interval recognition in detail. In particular, we enumerate and comment on the parameters which are determined for each experiment and the functions of each module used in our test program. Then we explain how the results of a series of experiments are summarized.

In Section 5.4, "Recognition Mistakes", we define two types of recognition errors which occur while recognizing intervals by simple correlation, false notes and missed notes. We analyze the situations which result in these two types of mistakes and suggest several refinements to avoid them, e.g. making spectral representations more accurate. On the other hand, it is explained that refinements in the model are limited by some practical reasons, e.g. the frequency accuracy is limited by the size of time windows. This means that the misrecognition can be reduced but cannot be eliminated completely.

In Section 5.5, "Efficiency and Stability of Recognition", the recognition of two-tone intervals, both harmonic and melodic, is analyzed. Since our model analyzes an unknown number of different pairs of correlated groups of partials and sorts out accidental correlations, the amount of computing and the extent to which the true recognition can be guaranteed are not known beforehand. In order to specify these characteristics, we attempt a special investigation and summarize its results. We conclude that the recognition of chords by recognizing intervals is already quite reliable but cannot be improved within the given simple correlation approach, so that some complements to the model are necessary.

In Section 5.6, "Testing the Decision Making Approach", we consider a model where each note of a chord is determined with taking into account several hypotheses obtained from the analysis of harmonic and different melodic intervals. We illustrate the advantage of the decision making approach and its contribution to the reliability of chord recognition with our series of computer experiments. We also show that the decision making rule may depend on the number of hypotheses processed and that the reliability of final recognition depends on the number of hypotheses as well.

In Section 5.7 "Testing the Structural Approach", we consider a model where the whole interval structure of a chord is recognized simultaneously. We analyze the performance of the model of chord recognition based on multi-correlation analysis of spectral data and on the use of the criterion of least complex data representation. In particular, we show that the limits of correct recognition are similar to the limits of human perception, and that the trends exhibited by the model correspond to that inherent in human recognition.

In Section 5.8, "Judging Computer Experiments", the main conclusions on the performance of the model are formulated.

5.2 Example of Chord Recognition

In order to illustrate the procedure of chord recognition and the way we analyze this procedure, consider two simple chords shown in Fig. 5.1a.

The power spectra of the chords are shown in Fig. 5.1b and their Boolean spectra are shown in Fig. 5.1c. The first chord, its power spectrum, and its Boolean spectrum are precisely the same as in Fig. 4.2, with the only difference that the frequency axis is vertical, in order to make the time axis horizontal.

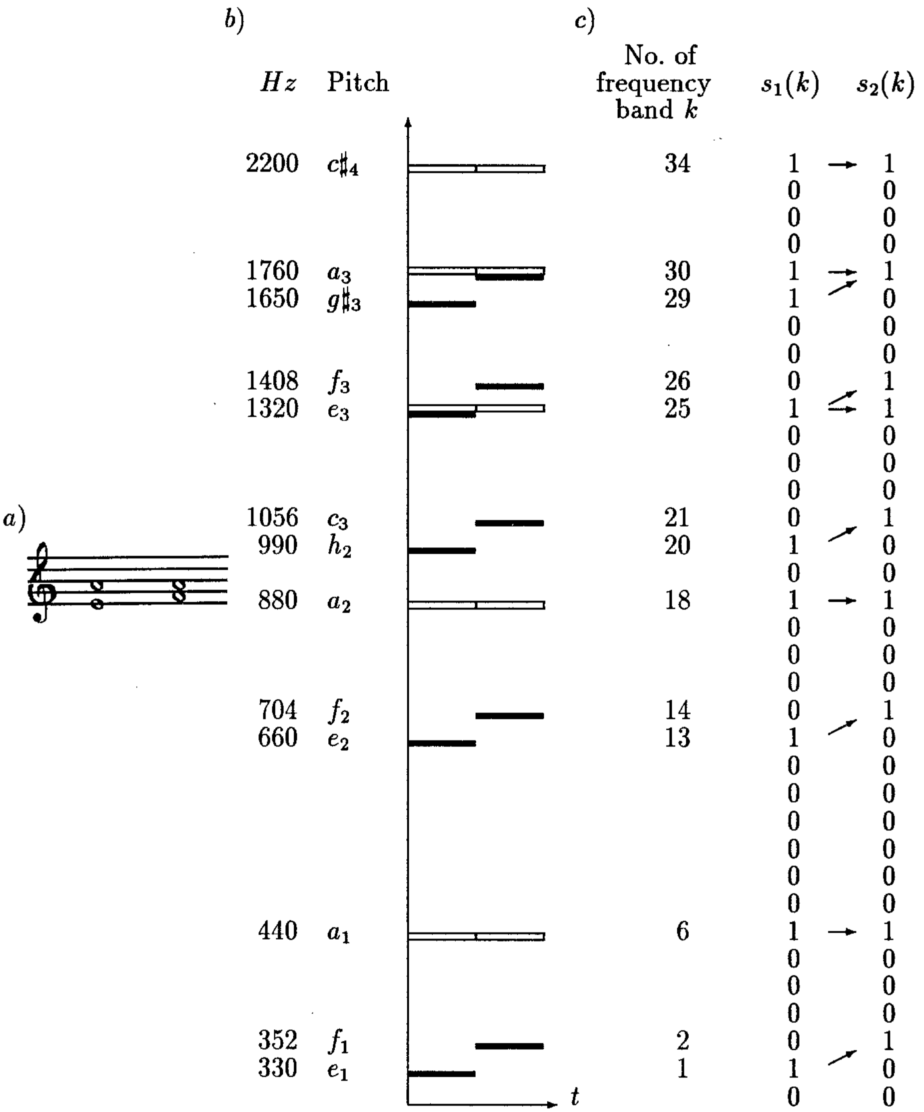

Figure 5.1: Spectral representations of chords

a) chords $(e_1; a_1)$ and $(f_1; a_1)$ in standard notation;

b) their audio spectra (with $\log_2$-scaled frequency axis) for harmonic voices with 5 successive partials (the partials of the lower and upper tones are shown by black or white rectangles, respectively);

c) the Boolean spectra of the chords under the frequency resolution within a semitone — the binary strings $s_n(k)$; the arrows show the parallel motion of partials.

Thus the nth chord is identified with the binary string

$$\{s_n(k)\},$$

where

n is the number of the given chord in the sequence of chords considered (in our example $n = 1, 2$),

k is the numbers of frequency bands in the spectrum considered (in our example $k = 1, \ldots, 34$),

$$s_n(k) = \begin{cases} 1 & \text{if there is a partial in the } k\text{th frequency band,} \\ 0 & \text{otherwise.} \end{cases}$$

We recognize melodic intervals between the nth chord and the $n+1$st chord by the peaks of correlation function

$$\begin{aligned} R_{n,n+1}(i) &= \sum_k s_n(k)s_{n+1}(k+i); \\ R_{n,n+1}(-i) &= \sum_k s_n(k)s_{n+1}(k-i) = \sum_k s_n(k+i)s_{n+1}(k). \end{aligned}$$

If this function has a peak at point i we suppose that the groups of partials correlated can contain tones which constitute the interval of i semitones, while positive i corresponding to ascending intervals and negative i corresponding to descending intervals. The subspectrum correlated is said to be a *generative group of partials* for the given melodic interval.

In other words, a melodic interval between the nth and the $n+1$st chord is determined by a correlated set of partials in the chord spectrum, and the *salience of melodic interval* of i semitones is identified with the value of the correlation function $R_{n,n+1}(i)$.

The correlation analysis of Boolean spectra of our two chords can be traced in Fig. 5.2. The strings $s_n(k)$ are shown with shifts in order to make the correlation illustrative. The value of correlation function $R(i)$ at point i is equal to the number of 1s coincided in correspondingly shifted strings. For example, $R_{1,2}(-1) = 0$ because there are no 1s coincided in the columns $s_1(k+1)$ and $s_2(k)$.

Table 5.1 displays the most salient melodic intervals not larger than the fifth, i.e. for the values $-7 \leq i \leq 7$. The generative group of partials of each melodic interval is denoted by the indexes of frequency bands. Although we don't use pitch in recognizing intervals, we *conventionally* denote the recognized intervals by notes, referring to the lowest partial of the correlated group of partials.

In Table 5.1 one can see the intervals $(a_1; f_1)$ and $(e_1; a_1)$, corresponding to the voice crossing. If the two voices had different timbre, i.e. different spectra,

Pitch	No. of frequency band k	$s_1(k+i)$	$s_2(k+i)$
$c\sharp_4$	34	1	1
		0 1	0 1
		0 · 1	0 · 1
		0 · · 1	0 · · 1
a_3	30	1 · · · 1	1 · · · 1
$g\sharp_3$	29	1 1 · · · 1	0 1 · · · 1
		0 1 1 · · · 1	0 · 1 · · · 1
		0 · 1 1 · · · 1	0 · · 1 · · · 1
f_3	26	0 · · 1 1 · · ·	1 · · · 1 · · ·
e_3	25	1 · · · 1 1 · ·	1 1 · · · 1 · ·
		0 1 · · · 1 1 ·	0 1 1 · · · 1 ·
		0 · 1 · · · 1 1	0 · 1 1 · · · 1
		0 · · 1 · · · 1	0 · · 1 1 · · ·
c_3	21	0 · · · 1 · · ·	1 · · · 1 1 · ·
h_2	20	1 · · · · 1 · ·	0 1 · · · 1 1 ·
		0 1 · · · · 1 ·	0 · 1 · · · 1 1
a_2	18	1 · 1 · · · · 1	1 · · 1 · · · 1
		0 1 · 1 · · · ·	0 1 · · 1 · · ·
		0 · 1 · 1 · · ·	0 · 1 · · 1 · ·
		0 · · 1 · 1 · ·	0 · · 1 · · 1 ·
f_2	14	0 · · · 1 · 1 ·	1 · · · 1 · · 1
e_2	13	1 · · · · 1 · 1	0 1 · · · 1 · ·
		0 1 · · · · 1 ·	0 · 1 · · · 1 ·
		0 · 1 · · · · 1	0 · · 1 · · · 1
		0 · · 1 · · · ·	0 · · · 1 · · ·
		0 · · · 1 · · ·	0 · · · · 1 · ·
		0 · · · · 1 · ·	0 · · · · · 1 ·
		0 · · · · · 1 ·	0 · · · · · · 1
a_1	6	1 · · · · · · 1	1 · · · · · · ·
		0 1 · · · · · ·	0 1 · · · · · ·
		0 · 1 · · · · ·	0 · 1 · · · · ·
		0 · · 1 · · · ·	0 · · 1 · · · ·
f_1	2	0 · · · 1 · · ·	1 · · · 1 · · ·
e_1	1	1 · · · · 1 · ·	0 1 · · · 1 · ·
		0 1 · · · · 1 ·	0 · 1 · · · 1 ·
	i	0 1 2 3 4 5 6 7	0 1 2 3 4 5 6 7
	$R_{n,n}(i)$	9 1 1 1 2 5 1 3	9 1 1 1 5 2 1 2
	$R_{1,2}(i)$		5 5 0 1 1 5 1 1
	$R_{1,2}(-i)$	5 0 1 1 6 1 1 2	

Figure 5.2: Correlation analysis of chord spectra (dots denote zeros)

these intervals would be less salient than the intervals corresponding to fluent leading of parts which are constituted by the same voice pattern.

Similarly, we recognize harmonic intervals in the nth chord by the peaks of autocorrelation function

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

The harmonic intervals are determined by correlated groups of partials in the chord spectrum, and the *salience of harmonic interval* of i semitones in the nth chord is identified with the value of the autocorrelation function $R_{n,n}(i)$.

Table 5.2 and Table 5.3 display the most salient harmonic intervals not larger than the fifth, i.e. for the values $0 \leq i \leq 7$, in the first and in the second chord, respectively. Note that the most salient harmonic interval in both tables is the interval of prime ($i = 0$), because the autocorrelation of the unshifted spectrum is always 100%.

The next step in our analysis is detecting the intervals which result from accidental correlations of partials. For example, such a coincidence arises in the descending interval of major third which is specified in Table 5.1 by $i = -4$ with the salience 6. The corresponding generative group of partials contains six partials: Five partials constitute the tone a_1 in the first chord and the tone f_1 in the second chord, and one accidentally joined partial. This accidental is the fifth harmonic of the tone e_1 in the first spectrum ($1650Hz$) matched to the third harmonic of the tone a_1 in the second spectrum ($1320Hz$).

In more complex cases, when the chord contains several tones and each tone is constituted by multiple partials, the correlation function may have peaks at the points which do not correspond to real intervals. This can result from accidental correlations of partials which belong to different tones, implying the recognition of these groups of accidentally correlated partials as tones.

As said in Section 5.1, in the first series of experiments we restrict our attention to the performance of the basic model, where harmonic and melodic intervals are recognized by simple correlation. Since we investigate the algorithm, we are interested in characterizing the situations when correlated groups of partials do not correspond to real intervals. In our case of synthesized spectra, such situations are recognized by comparing the correlated groups of partials with the spectral *standard* used in the generation of chord spectra. We characterize a given correlated group by the *number of successive partials* of the standard inherent in the given group.

In our example, the voices have five successive harmonics with the frequency ratio $1 : 2 : 3 : 4 : 5$. According to Example 5 from Section 4.2, the voice standard is determined by the frequency bands indexed by $0, 12, 19, 24, 28$. In Table 5.1 one can find the group of six partials corresponding to the descending interval of major third ($i = -4$) which contains the standard pattern whose partials are indexed by $6, 18, 25, 30, 34$ (that is $0, 12, 19, 24, 28$ translated). Con-

Table 5.1: Most salient melodic intervals between two chords

Correlation $R_{1,2}(i)$	Interval i	Generative group of partials	Number of successive harmonics	Notation of the interval
6	−4	6 18 25 29 30 34	5	$(a_1; f_1)$
5	0	6 18 25 30 34	5	$(a_1; a_1)$
5	1	1 13 20 25 29	5	$(e_1; f_1)$
5	5	1 13 20 25 29	5	$(e_1; a_1)$
2	−7	13 25	2	$(e_2; a_1)$

Table 5.2: Most salient harmonic intervals in the first chord

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

Table 5.3: Most salient harmonic intervals in the second chord

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

sequently, the correlated group of partials contains all the five partials of the standard, confirming the right recognition of true interval.

If the number of successive partials was less than five, this would mean that the correlated group of partials doesn't contain the tone pattern used in the chord spectrum generation. In our analysis such a group of partials is immediately recognized as accidental. The insufficient number of successive partials is a simple but efficient criterion for recognizing accidental correlations for testing the performance of the algorithm of interval recognition.

The results of comparing each generative pattern with the standard are displayed in the fourth column of Tables 5.1–5.3. If an interval is rather salient (i.e. if it is characterized by a considerable correlation, meaning that the correlated group of partials is numerous), but the generative group of partials doesn't match the standard, we conclude that we have revealed a *false* interval.

The next step in our analysis is testing the two approaches to reducing the number of mistakes which have been outlined in Section 5.1, decision making approach and structural approach.

The decision making approach is based on the observation that a true tone belongs usually to several harmonic and melodic intervals. If these intervals are identified with certain conventional notes, these notes are recognized several times, whereas the notes corresponding to accidental intervals are most likely recognized only once. Since every chord can be recognized separately by harmonic intervals, or by melodic intervals, being confronted to another chord, the final decision on the chord's compound can be made with taking into account the frequency of the notes in the preliminary hypotheses.

In our model we have considered three hypotheses on each chord from the given progression. The first hypothesis that a given generative group of partials contains a true tone pattern is obtained from the analysis of harmonic intervals of the chord. Another hypothesis is obtained from the analysis of melodic intervals between the given and preceding chord. The third hypotheses is obtained from the analysis of melodic intervals between the given and succeeding chord.

One can consider more hypotheses, analyzing melodic intervals between the given chord and remote chords. For this purpose one should confront the given chord spectrum to spectra of remote chords, as if tracking latent voice leading. However, we restrict our attention to the three hypotheses mentioned, derived from the analysis of the local context.

As shown below, the choice of a decision making procedure depends on the number of hypotheses considered. If we consider all possible hypotheses (as many as there are chords in the analyzed progression), a *majority rule* can be recommended (a tone is accepted if it is backed up by at least a half of the hypotheses). In our case with three hypotheses only, the best results are obtained by their *logical disjunction* (a tone is accepted if it is backed up by at least one hypothesis). c In our simple example, the two decision making rules are equivalent, since there are no mistakes in recognizing tones from salient intervals. However, as seen below, in more complex cases the misrecognition of intervals is quite frequent, and the decision making approach is rather useful.

The structural approach is based on the observation that the structure of partial groups which correspond to real tones is repeated regularly, constituting a stable spectral pattern. On the contrary, accidentally correlated groups of partials which do not correspond to real tones have a random spectral structure. Therefore, in order to recognize an accidental correlation, one has to recognize the irregularity of the structure of the group of partials correlated. This can be done by further correlation analysis applied to different generative groups of partials. If a generative group doesn't correlate with other generative groups, meaning that their structures have little in common, it is almost certain that the given generative group of partials is accidental and then it is rejected. This is the idea of multicorrelation analysis from the previous chapter which filters out accidentally correlated groups of partials. After the whole interval structure of a chord is determined, the groups of partials multicorrelated can be associated with conventional pitches.

The generative groups of partials for our simple example are shown in the third column of Tables 5.1–5.3. One can easily see that all the generative groups of partials with a sufficient salience have the same underlying subspectrum with the partials indexed by $0, 12, 19, 24, 28$. This means that this structure predominates in the generation of the given chords, and after identifying the pitches of its different appearances, one obtains the two chords from our example.

5.3 Testing the Simple Correlation Approach

In order to analyze the procedure of chord recognition, we have performed a series of computer experiments with synthesized spectra of chords.

These spectra have been computed according to certain rules and assumptions by the **first program module**. This module performs the data generation according to some initial parameters. For each experiment the user determines:

1. *Progression of chords* coded in letter names of notes. For example, $(A, e, c\sharp_1, a_1)$ denotes *la major* four-part chord, where capital letters mean the great octave, small letters mean small octave, indexed letters mean the octaves higher than the middle *do*, e.g. a_1 is the *la* of the one-line octave.

2. *Voice type.* We have used *harmonic*, with integer ratio of partial frequencies $1 : 2 : 3 : \ldots$, and *inharmonic*, with the frequency ratio $1 : \sqrt{2} : 2 : \sqrt{2^3} : \ldots$.

3. *Number of partials per voice.* We have used 5, 10, or 16 partials per voice.

4. *Accuracy of spectral representation*, or *frequency resolution.* We have used the accuracy to within 1, 1/2, 1/3, and 1/6 semitone.

5. *Maximal intervals considered.* This parameter is aimed at

   (a) providing means to avoid the octave autocorrelation since odd partials of a harmonic tone can be misrecognized as an independent tone; for this purpose the maximal intervals considered are restricted to 12 semitones;

   (b) restricting the correlation analysis to a limited diapason where the voice spectra are not expected to differ considerably; this effect is known in musical instruments, e.g. the clarinet voice is quite different in low and high registers not only due to the difference in partial intensities but also because of different number of partials.

   In our experiments we have used the restriction up to 12 semitones, and no restriction.

For the example from the previous section, the input parameters are written at the top of Table 5.4. Then for each chord its Boolean spectrum is computed in the form of binary string, similarly to that shown in Fig. 5.1c. When the frequency resolution is more accurate than one semitone, the corresponding binary strings become longer, and the 1s become more rare. This results in fewer accidental correlations of partials from different voices.

The **second program module** performs correlation analysis of Boolean spectra. It computes the autocorrelation function for a given chord spectrum, and the correlation function for the spectra of the chord and its predecessor. The result of this procedure is displayed in Table 5.4. Then the program sorts the harmonic and melodic intervals with respect to the correlation values, putting more salient intervals at the top of the corresponding list like in Tables 5.1–5.3 5.4.

The **third program module** processes the most salient intervals, comparing them with the standard voice spectrum which has been used in the spectrum generation. In our example, this standard is determined by the partial frequency ratio $1 : 2 : 3 : 4 : 5$. This comparisons are aimed at estimating efficiency and stability of recognition, and for this purpose certain characteristics of each program run are stored in the memory.

Table 5.4: Correlations of chord spectra

Number of chords = 2
Number of partials per voice = 5
Ratio of partial frequencies = 1 : 2 : 3 : 4 : 5
Accuracy of representation = 1 semitone
Maximal intervals considered, up to = 12 semitones
Input chords: $(e_1; a_1)$, $(f_1; a_1)$

Number of current chord n	Interval in semitones i	Correlation: For harmonic interval $R_{n,n}(i)$	For melodic interval from preceding chord: Descending $R_{n-1,n}(-i)$	Ascending $R_{n-1,n}(i)$
1	0	9	0	0
	1	1	0	0
	2	1	0	0
	3	0	0	0
	4	2	0	0
	5	5	0	0
	6	0	0	0
	7	3	0	0
	8	0	0	0
	9	2	0	0
	10	1	0	0
	11	1	0	0
2	0	9	5	5
	1	1	0	5
	2	0	1	0
	3	1	1	1
	4	5	6	1
	5	2	1	5
	6	0	1	1
	7	2	2	1
	8	3	2	3
	9	2	2	1
	10	0	0	1
	11	1	3	0

For each chord and each type of interval (harmonic interval, ascending melodic interval, and descending melodic intervals), it is calculated how many intervals from the most correlated ones should be processed in order to reveal all the tones of a given chord. Obviously, the desired information about true harmonic and melodic intervals is represented by the most salient intervals, corresponding to the peak values of correlation function. Therefore, we scan over the intervals, beginning from the most correlated ones.

The minimal set of the most salient intervals of the given type, which contains all true tone patterns of a given chord, is said to be the *sufficient set (of interval patterns)*. The number of these interval patterns conventionally characterizes the *efficiency* of the recognition procedure, corresponding to the amount of necessary computing.

The sufficient set of patterns may contain accidental patterns. The *number of accidental patterns in the sufficient set* also characterizes the efficiency of recognition, corresponding to the amount of "useless" computing. In our simple example there are no accidental patterns, but only accidental partials joined to true patterns. However, in more complex cases salient accidental patterns emerge frequently.

In order to estimate the *stability* of the recognition procedure, we determine the *minimal number of successive standard partials* which may be inherent in true patterns from the sufficient set. If this minimal number is small then all accidental patterns are "very irregular" and, consequently, well distinguishable from true patterns. On the contrary, if this minimal number is close to the number of partials of the standard, then false tones are not well distinguishable from true patterns.

Obviously, the lower is this minimal number of successive standard partials, the larger is the difference between true and accidental patterns, implying the stability of the recognition with respect to possible distortions of data. Certainly, the minimal number of successive standard partials should be regarded with the reference to the number of partials in the standard. For instance, if the standard contains 10 partials then the threshold value 4 is rather small, but if it contains only 5 partials then the same threshold value 4 should be regarded as large, meaning low stability of true recognition.

In order to provide for such a relative comparison, we always display the *maximal number of successive standard partials* inherent in true patterns from the sufficient set.

Note that if the sufficient set is large, the minimal number of successive standard partials is usually large as well (the more patterns, the more the risk of accidentals). From this we conclude that efficiency and stability are interdependent in our model.

The results of such an analysis of the recognition procedure for our simple example is displayed in Table 5.5. In our example, there are no accidental

Table 5.5: Specifications of recognition procedure

Input chords	Sufficient number of patterns/ Number of accidentals in the sufficient set of patterns/ Minimal–maximal number of successive standard partials inherent in true patterns from the sufficient set		
	By harmonic intervals	By melodic intervals: From the preceding chord	By melodic intervals: To the succeeding chord
1. $(e_1; a_1)$	2/0/1–5	—	3/0/1–5
2. $(f_1; a_1)$	2/0/1–5	2/0/1–5	—

partials in the sufficient set. Consequently, the minimal number of successive standard partials is extremely small, being equal to 1, which means high stability. If there was an accidental pattern in the sufficient set, e.g. with 2 successive partials from the spectral standard, then the minimal number of successive standard partials would rise to the value 3.

In a longer progression of chords these estimates are processed as some statistical characteristics and some conclusions on the reliability and stability of the recognition are made for the given run of the program. This is the task for the **fourth program module.**

In a series of experiments where the initial conditions are different, the estimates obtained for each experiment are stored in the memory. Then the results are generalized over all these experiments as well, and general conclusions about the dependence of reliability and stability of recognition on the parameters of the experiments are made. This is the task for the **fifth program module.**

5.4 Recognition Mistakes

The algorithm described commits two types of errors:

- *false (wrong) notes*, and
- *missed (omitted) notes.*

According to such a classification of mistakes, a misplaced note is counted as two mistakes. For example, if the chord (C, E, G) is recognized as (C, D, G), it has one false note D and one missed note E.

Generally speaking, the number of mistakes is the sum of missed and false notes which is the number of notes in the symmetrical difference between

Table 5.6: Specifications of recognition procedure restricted to 12 semitones

Input chords	Sufficient number of patterns / Number of accidentals in the sufficient set of patterns / Minimal–maximal number of successive standard partials inherent in true patterns from the sufficient set		
	By harmonic intervals	By melodic intervals	
		From preceding chord	To succeeding chord
1. $(G; f; c_1; g_1)^*$	38/34/11–16 $-g_1$	—	74/59/11–16 $-g_1$
2. $(C; e; c_1; g_1)^*$	38/35/11–16 $-g_1$	74/63/11–16 $-g_1$	—

* the chords misrecognized in all of the three hypotheses
– missed tone in the corresponding hypothesis

Table 5.7: Specifications of recognition procedure not restricted

Input chords	Sufficient number of patterns / Number of accidentals in the sufficient set of patterns / Minimal–maximal number of successive standard partials inherent in true patterns from the sufficient set		
	By harmonic intervals	By melodic intervals	
		From preceding chord	To succeeding chord
1. $(G; f; c_1; g_1)$	7/3/11–16	—	16/3/11–16
2. $(C; e; c_1; g_1)$	11/6/11–16	13/6/11–16	—

Figure 5.3: Two hardly recognizable chords

"truth" and "recognition". Thus in the given example

$$\begin{aligned}\text{Mistakes} &= (C,E,G)\,\triangle\,(C,D,G)\\ &= [(C,E,G)\cup(C,D,G)]\setminus[(C,D,G)\cap(C,D,G)]\\ &= (C,D,E,G)\setminus(C,G)\\ &= (D,E).\end{aligned}$$

A false note emerges from accidental correlation of partials of irrelevant voices. As mentioned in Section 5.1, accidentally correlated groups of partials have random spectral structure and can be recognized as irregular by comparing them with other generative groups of partials. Being quite efficient, this method nevertheless doesn't guarantee the 100%-reliability.

The case of missed notes is more difficult. In order to illustrate it, consider two four-part chords shown in Fig. 5.3. Suppose that each voice consists of 16 harmonics, and that the accuracy of representation is within 1/3 semitone.

In Tables 5.6 and 5.7 one can see the specifications of two experiments on the recognition of these chords differing in the restriction on the maximal size of intervals considered, up to 12 semitones and with no restriction, respectively.

In the experiment illustrated by Table 5.6 the correct recognition of the two chords is impossible. Consequently, no set of patterns is sufficient. In this case our program scans over all possible intervals, and instead of sufficient number of patterns it prints the total number of patterns scanned (38 for harmonic intervals, and 74 for both types of melodic intervals). Therefore, the number of accidental patterns in the sufficient set of patterns is equal to the total number of accidental patterns (34, or 35 for harmonic intervals, and 63 or 59 for melodic intervals).

While processing an accidental pattern, our program counts the number of successive standard partials in this accidental pattern, which is usually much less than their number in a true tone pattern. In order to find the threshold for distinguishing accidentals, the program determines the minimal number of successive standard partials in true tones which is by one greater than the number of successive standard partials in the accidental patterns processed. In our case this threshold is as high as 11 (with the reference to 16 partials in the voice standard).

Since in the given experiment the harmonic intervals are restricted to 12 semitones, the tone g_1 of the first chord $(G; f; c_1; g_1)$ can be recognized by the only true interval—the fifth $(c_1; g_1)$. Indeed, other intervals which contain g_1 are larger than 12 semitones and cannot be analyzed. However, the correlated groups of partials corresponding to the fifth also contain the partials associated with the interval $(f; c_1)$. This means that the harmonic interval $(f; c_1)$ *masks* the upper parallel interval $(c_1; g_1)$, and the tone g_1 becomes indistinguishable with respect to the only interval of fifth.

In the second chord the note g_1 is also missed when recognized by harmonic intervals. The only true harmonic interval smaller than 12 semitones which contains g_1 is again the interval of fifth $(c_1; g_1)$. However, the correlated groups of partials corresponding to the fifth contain accidentally correlated second overtone of C and first partial of g so that the interval of fifth is misrecognized as $(c; g)$ which "masks" the true interval $(c_1; g_1)$.

The tones g_1 are also missed when the chords are recognized by melodic intervals between them. The only true melodic intervals smaller than 12 semitones which contain g_1 are $(g_1; g_1)$, $(g_1; c_1)$, and $(c_1; g_1)$. However, these intervals are also masked by accidental lower interval $(g; g)$, and true lower parallel intervals $(G; C)$ and $(f; c_1)$, respectively.

In the second experiment illustrated by Table 5.7, with no restriction on the intervals considered, the situation is more favorable. Each tone can be put into correlation with the lowest note of the chord, and thus the corresponding interval cannot be masked by some lower interval, since there are no lower tones. This observation is valid for the recognition by harmonic intervals and by melodic intervals as well.

In spite of the better recognition with no restriction on intervals considered, one should take into account that in reality this restriction is important. It prevents misrecognizing odd partials of a harmonic voice as an independent voice (which results from the octave autocorrelation of harmonic voices). Moreover, the spectra of acoustical sounds change greatly with pitch, which reduces the correlation of tones from different registers even of the same musical instrument. Since low tones are usually rich in harmonics and high tones, on the contrary, quite poor, the audio diapason should be processed in several registers where the tones have similar spectra.

5.5 Efficiency and Stability of Recognition

In order to test our model, we have performed a series of computer experiments on chord recognition by recognizing harmonic and melodic intervals. For the experiments we have chosen the 130th four-part chorale from J.S.Bach's *371 Four-Part Chorales* shown in Fig. 5.4 The chorale has been considered as a sequence of 24 chords, corresponding to the harmonic verticals.

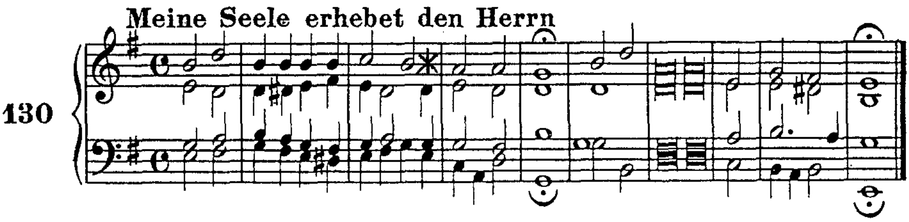

Figure 5.4: The 130th chorale from J.S.Bach's *371 Four-Part Chorales* (the asterisk corresponds to the note missed in recognition)

The Boolean spectra of the chords have been computed and analyzed under various conditions. The experiments have differed in the following parameters listed at the beginning of Section 5.3:

- Voice type (harmonic, with partial frequency ratio $1 : 2 : 3 : \ldots$, or inharmonic, with partial frequency ratio $1 : \sqrt{2} : 2 : \sqrt{2^3} : \ldots$);
- number of partials per voice (5, 10, or 16);
- accuracy of spectral representation (within 1, 1/2, 1/3, and 1/6 semitone);
- restriction on maximal intervals considered (up to 12 semitones, or no restriction).

Having tried all combinations of these parameters, we have performed totally 48 experiments.

A typical computer output for one of these 48 experiments is displayed in Table 5.8. At the bottom of this table one can see some general characteristics of the given experiment. The maximal values characterize the worst recognition situations, in particular, when the correct recognition of all true patterns is impossible. The maximal values for recognized chords characterize the amount of computing which is necessary to correctly recognize all recognizable true patterns. The average values for recognized chords characterize the average amount of computing for correctly recognizing all recognizable true patterns. Finally, the average values characterize the average amount of computing for processing each chord.

As said in Section 5.3, the results of all the 48 experiments have been summarized by the fifth program module in order to estimate the efficiency and stability of the recognition procedure. The characteristics which are shown

Table 5.8: Specifications of recognition procedure for one experiment on chord recognition
Type of voice: Harmonic
Number of partials per voice: 16
Accuracy of spectral representation: Within 1/2 semitone
Restriction on maximal intervals considered: Up to 12 semitones

Input chords[1]	Sufficient number of patterns / Number of accidentals in the sufficient set of patterns / Minimal–maximal number of successive standard partials inherent in true patterns from the sufficient set		
	By harmonic intervals	By melodic intervals	
		From preceding chord	To succeeding chord
1.$(e; g; e_1; h_1)$	3/0/1–16	—	7/1/9–16
2.$(f\sharp; a; d_1; d_2)$	24/19/10–16 $-d_2$	11/1/4–16	46/33/10–16 $-d_2$
3.$(g; h; d_1; h_1)$	10/5/11–16	8/0/1–16	5/0/1–16
4.$(f\sharp; a; d\sharp_1; h_1)$	8/3/13–16	5/0/1–16	9/1/13–16
5.$(e; g; e_1; h_1)$	3/0/1–16	5/0/1–16	9/1/9–16
6.$(d\sharp; f\sharp; f\sharp_1; h_1)$	3/0/1–16	5/1/10–16	6/2/10–16
7.$(e; g; e_1; c_2)$	4/1/7–16	8/1/7–16	9/0/1–16
8.$(f\sharp; a; d_1; c_2)$	6/2/10–16	5/0/1–16	5/0/1–16
9.$(g; a; d_1; h_1)$	6/1/6–16	5/0/1–16	10/1/4–16
10.$(e; g; d_1; h_1)^*$	24/19/11–16 $-h_1$	46/34/11–16 $-h_1$	46/31/11–16 $-h_1$
11.$(c; g; e_1; a_1)$	24/20/12–16 $-a_1$	8/2/11–16	46/32/11–16 $-a_1$
12.$(A; g; e_1; a_1)$	5/2/11–16	46/36/12–16 $-e_1$	7/2/11–16

13.($d; f\sharp; d_1; a_1$)	4/0/1–16	9/1/9–16	7/0/1–16
14.($G; h; d_1; g_1$)	6/3/11–16	10/4/11–16	46/38/11–16 $-G$
15.($g; g; d_1; h_1$)	5/2/11–16	8/3/11–16	8/1/11–16
16.($H; g; d_1; d_2$)	24/20/10–16 $-d_2$	5/1/9–16	46/36/10–16 $-d_2$
17.($d; f\sharp; d_1; a_1$)	4/0/1–16	46/32/9–16 $-a_1$	7/0/1–16
18.($d; f\sharp; d_1; a_1$)	4/0/1–16	6/0/1–16	46/35/10–16 $-a_1$
19.($c; a; e_1; e_1$)	2/0/1–16	7/1/9–16	3/0/1–16
20.($H; h; e_1; g_1$)	3/0/1–16	6/0/1–16	9/1/7–16
21.($A; h; e_1; g_1$)	5/2/11–16	6/1/11–16	7/0/1–16
22.($H; h; d\sharp_1; f\sharp_1$)	4/1/8–16	6/0/1–16	5/0/1–16
23.($H; a; d\sharp_1; f\sharp_1$)	3/0/1–16	8/3/11–16	46/31/11–16 $-f\sharp_1$
24.($E; g; h; e_1$)	7/4/11–16	8/2/11–16	—
Maximal values	24/20 Fig. 5.5 / 13–16	46/36/12–16	46/38/13–16
Maximal values for recognized chords	10/5 / 13–16 Fig. 5.7	11/4/11–16	10/2/13–16
Average values for recognized chords	5/1 Fig. 5.6 / 6–16	7/1/6–16	7/1/5–16
Average values	8/4 / 7–16	12/5/7–16	19/11/7–16
Number of recognized chords	20	20	16

[1] h denotes note $si\natural$

* the chords misrecognized in three hypotheses

– missed tone in the corresponding hypothesis in frames are displayed later in diagrams which trace their behavior over all of the 48 experiments.

Now let us outline the main conclusions from our experiments.

The recognition of harmonic voices with 5 partials is independent of the accuracy of representation. This is explained by the precise coincidence of the first 5 harmonics with the degrees of the tempered scale, whence no further separation of partials is possible by refining the spectral representation. This implies that the correlation analysis of spectral data reveals the same intervals under any accuracy of spectral representation.

The harmonic voices with multiple partials are better recognizable when the spectral representation is more accurate. This is explained by the fact that under a poor frequency resolution a large number of partials in voice spectra makes the chord spectrum too dense, resulting in many accidental correlations and thus making difficult distinguishing true patterns from accidentals. Therefore, a refinement of the representation improves the separation of partials of irrelevant voices, implying better recognizability of voices.

For inharmonic voices the recognition is more reliable in case of more accurate frequency representation, both for 5 partials per voice, and for 10 or 16 partials per voice. This is explained by the fact that the partials of inharmonic voices do not fall at the centers of frequency bands which are adjusted to the tempered scale (even in case of 5 partials, unlike harmonic voices). This implies that two partials from two different inharmonic voices, which are indistinguishable under a poor frequency resolution, can be separated in a more accurate representation. That is why accidental correlations of partials of irrelevant voices, which result in accidental interval patterns, occur in an accurate representation more seldom, making chord recognition simpler.

The experiments show that the recognition efficiency and stability depend on voice type, accuracy of representation, and restriction on maximal intervals considered. For the chord recognition by harmonic intervals, these dependencies are displayed by graphs in Fig. 5.5–5.7 where one can trace the behavior of the characteristics shown in frames in Table 5.8 versus representation accuracy and number of partials per voice.

Analogous characteristics for chord recognition by melodic intervals are given in the third and fourth columns in Table 5.8. They have similar properties, so we do not provide figures for them.

The numbers $1, 2, 3, 4, 5, 6$ under the axes of spectral resolution denote the accuracy of representation within $1, 1/2, \ldots, 1/6$ semitone, respectively. (Since in computer experiments we have used only $1, 1/2, 1/3$, and $1/6$, the points corresponding to the accuracy $1/4$ and $1/5$ are linearly extrapolated.) Each diagram displays three cuts, associated with the voices with $5, 10$, or 16 partials. The vertical axis is graduated by 5 units marked by dots.

In Fig. 5.5a and Fig. 5.5c one can see that under the restriction on maximal

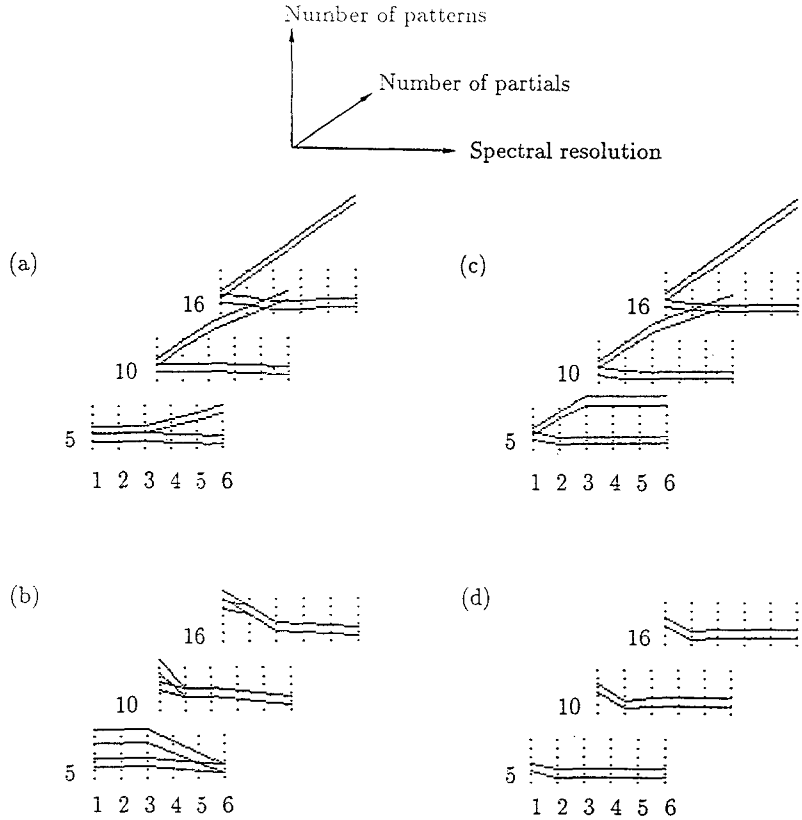

Figure 5.5: Maximal values of the sufficient number of patterns and maximal number of accidentals in the sufficient set of patterns
(upper pair of curves for the whole experiment, the lower pair of chords for only correctly recognized chords in the experiment)

(a) for harmonic voices and maximal intervals considered up to 12 semitones;

(b) for harmonic voices and no restriction on maximal intervals considered;

(c) for inharmonic voices and maximal intervals considered up to 12 semitones;

(d) for inharmonic voices and no restriction on maximal intervals considered.

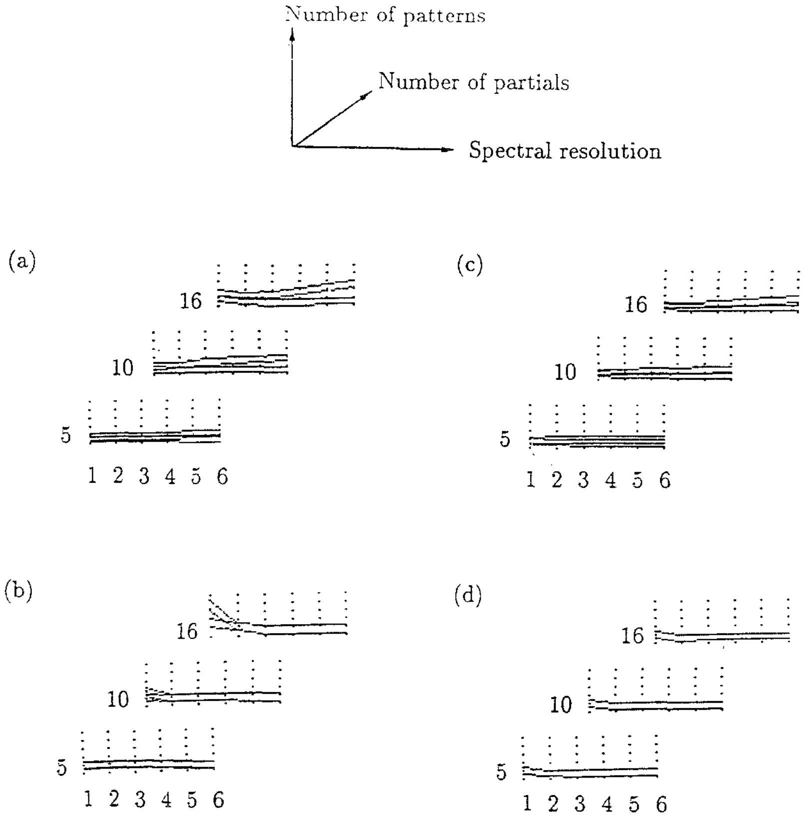

Figure 5.6: Average value of the sufficient number of patterns and average number of accidentals in the sufficient set of patterns
(upper pair of curves for the whole experiment, the lower pair of chords for only correctly recognized chords in the experiment)

(a) for harmonic voices and maximal intervals considered up to 12 semitones;

(b) for harmonic voices and no restriction on maximal intervals considered;

(c) for inharmonic voices and maximal intervals considered up to 12 semitones;

(d) for inharmonic voices and no restriction on maximal intervals considered.

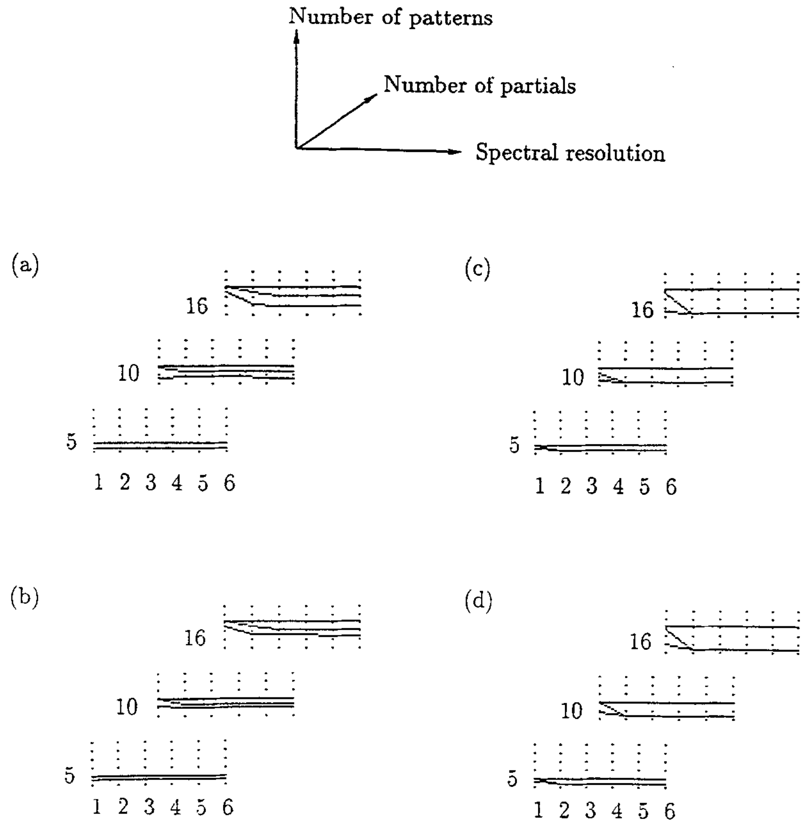

Figure 5.7: Maximal and average values of the limits of the number of successive standard partials inherent in true patterns from the sufficient set (upper curve is the maximal value of successive standard partials in patterns of sufficient set, the middle curve is the maximum value of minimal number of successive standard partials, the lower curve is the average value of minimal number of successive standard partials)

(a) for harmonic voices and maximal intervals considered up to 12 semitones;

(b) for harmonic voices and no restriction on maximal intervals considered;

(c) for inharmonic voices and maximal intervals considered up to 12 semitones;

(d) for inharmonic voices and no restriction on maximal intervals considered.

intervals considered, the maximal sufficient number of patterns grows with a refinement in the representation accuracy. Indeed, in this case the 100%-correct recognition is impossible, and therefore no set of patterns is sufficient. In our analysis this means that we exhaust the set of possible patterns which is obviously larger for more accurate representations (for example, if the accuracy is within 1 semitone we have 12 intervals per octave, yet if the accuracy is within 1/2 semitone this number is twice greater, implying more interval patterns). The increase in the sufficient number of patterns implies that the number of accidental patterns increases as well. For correctly recognized chords (lower pair of curves in the diagrams), the curves decrease while refining the representation accuracy, because, as mentioned above, salient accidental patterns occur more seldom.

In Fig. 5.5b and Fig. 5.5d one can see that under no restriction on maximal intervals considered the maximal sufficient number of patterns decreases while refining the representation accuracy, i.e. the refinement of representation improves the recognition efficiency. The coincidence of the two pairs of curves means that the 100%-correct recognition is obtained.

Fig. 5.6 illustrates the fact that the average values of sufficient number of patterns and the number of accidental patterns in the sufficient set are much smaller then the maximal values shown in Fig. 5.5. The upper pair of curves in Fig. 5.6a and Fig. 5.6c grows much slower than that in Fig. 5.5a and Fig. 5.5c. This means that the finer is the representation accuracy, the fewer are the cases of misrecognition. In other words, the recognition becomes more reliable, on average requiring fewer patterns to be processed.

The stability of the recognition procedure is illustrated by Fig. 5.7. The distance between the two upper curves can be interpreted as a gap between true and wrong patterns. The larger is this gap, the more seldom an accidental pattern is recognized as a true one. The distance between the two lower curves corresponds to the difference between the greatest similarity of false and true patterns on the one hand, and the average estimation of this similarity on the other hand.

Therefore, the greater the first distance and the smaller the second distance are, the more stable is the recognition procedure. From Fig. 5.7b and Fig. 5.7d one can conclude that the recognition of inharmonic voices is much more stable than that of harmonic voices illustrated by Fig. 5.7a and Fig. 5.7c, and that the recognition stability for inharmonic voices is almost independent of the restriction on maximal intervals considered, which is nevertheless a little better in case when there is no restriction.

5.6 Testing the Decision Making Approach

As mentioned in Section 5.4, the algorithm described commits two types of errors, false notes and missed notes.

False notes emerge under an insufficient accuracy of spectral representation (within a semitone) when accidental correlations are quite probable. For example, when the accuracy of representation is within one semitone, the harmonics after the 12th are so dense in the spectrum (Fig. 3.1) that they fill all successive frequency bands in, implying the correlation of every two groups of high partials.

Therefore, one can make spectral representation more accurate, "rarefying" the spectrum. Then true and false patterns are easier distinguishable, which results in decreasing in the number of false notes in the recognition of chords.

Another cause of false notes is a low salience of voice spectra, when the number of partials per voice is small, which in our experiments corresponds to 5 partials. Indeed, accidentally correlated small groups of partials are more probable than accidentally correlated large groups. Therefore, accidentally correlated groups of partials are better distinguishable from true patterns in case when voice spectral patterns are larger. However, this is true for a sufficiently accurate representation.

On the other hand, refining the accuracy of spectral representation can help in distinguishing true voice patterns with low salience from accidentals. This is caused by the fact that a replication of the same combination, say, of 5 accidental partials is less probable in a rarefied spectrum. For instance, under representation accuracy within one semitone such an accidental correlation requires a coincidence in 5 from 29 frequency bands, yet under accuracy within 1/2 semitone it requires a coincidence in 5 of 57 frequency bands, which is less probable. Such a frequency explanation is most evident for inharmonic voices, since at the standard tempered scale the partials of different inharmonic tones do not coincide with each other and therefore are separable by refining the accuracy of spectral representation.

Missed notes emerge when a true interval pattern is masked by a lower parallel interval (see Section 5.4). Therefore, to avoid missing notes one should avoid lower "maskers". For this purpose the intervals with the lowest tone of a chord can be processed, corresponding to the condition of no restriction on maximal intervals considered.

However, the two ways of eliminating recognition mistakes are efficient only in the model with synthesized data. As already mentioned, the lack of restriction on maximal intervals considered can result in recognizing odd partials of a harmonic voice as an independent pattern. On the other hand, the accuracy of spectral representation of acoustical signal is also limited by the time resolution: The smaller are time windows, the poorer is the frequency accu-

Table 5.9: Preliminary and final results of chord recognition

Voice type	Maximal number of misrecognized notes in a preliminary hypothesis (over three hypotheses) / Number of misrecognized notes after decision making by a majority rule / Number of misrecognized notes after decision making by logical disjunction of three hypotheses					
	For voices with 5 partials and accuracy of representation		For voices with 10 partials and accuracy of representation		For voices with 16 partials and accuracy of representation	
	1 semitone	1/2 semitone	1 semitone	1/2 semitone	1 semitone	1/2 semitone
Harmonic	8/5/2	8/5/2	10/8/4	7/3/1	18/16/11	7/4/1
Inharmonic	4/3/1	4/2/0	4/3/1	4/2/0	4/3/1	4/2/1

racy. Moreover, taking into account various disturbances of natural sounds, our method is expected to fail under fine accuracy of spectral representation, since under this condition even small variances in spectra can break the correlation of matched tones.

To a certain extent, these difficulties can be overcome by using a decision making procedure which unites the hypotheses about the chord to be recognized. In our experiments each chord is recognized three times:

- by harmonic intervals within the given chord;
- by melodic intervals between the given chord and its predecessor;
- by melodic intervals between the given chord and its successor.

The only exceptions are the first chord and the last chord, which are recognized in two appearances.

Let us illustrate the effect of such a decision making procedure with an example from our series of experiments. Consider the experiments under the worst conditions, when the accuracy of spectral representation is within 1 or 1/2 semitone and when maximal intervals considered are restricted to 12 semitones.

The most evident way to make the final decision is to apply a *majority rule*, i.e. to accept tones backed up by at least two of the three hypotheses. However, owing to our method of designating notes by the lowest partial of a spectral pattern, missing notes occur much more often than false notes, and better results can be obtained by *logical disjunction of the hypotheses*, i.e. by accepting the tones backed up by at least one hypothesis.

Table 5.9 displays the final recognition results for the 24 chords of the J.S.Bach chorale under these worst conditions. The numbers in the table should be understood in the following way. For example, for the experiment illustrated by Table 5.8 the recognition by harmonic intervals gives 4 mistakes, by melodic intervals from preceding chords—3 mistakes, and by melodic intervals to succeeding chords—7 mistakes. This means that the last preliminary hypotheses gives the maximal number of mistakes 7. After having applied a majority rule to each note recognized, we reduce the number of mistakes to 4 (missed d_2 from the second chord, missed h_1 from the 10th chord, missed a_1 from the 11th chord, and missed d_2 from the 16th chord), and after having applied the logical disjunction, we reduce the number of mistakes is 1 (missed h_1 from the 10th chord, the only mistake inherent in the three hypotheses simultaneously). These indicators for this experiment are displayed in the form 7/4/1 at the right end of the first line of Table 5.9.

One can see that the decision making procedures reduce the number of misrecognized notes. Thus under accuracy of spectral representation within 1/2 semitone the number of mistakes in the worst case is reduced from 8 to

2. Since the four-part chorale represented by 24 harmonic verticals contains $4 \times 24 = 96$, (yet the parts are superimposed two times), 8 mistakes correspond approximately to reliability of recognition 91%, and 2 mistakes to almost 98%. This means that the decision making approach improves the reliability of chord recognition from 91% to 98% (in the worst case).

The recognition reliability can be improved further by considering more hypotheses, derived from correlation analysis of pairs of remote chords. Let us show this with an example of recognizing note h_1 ($si_1\natural$) from the 10th chord in the experiment illustrated by Table 5.8 (marked in Fig. 5.4 and in Table 5.8 by the asterisk).

The failure in recognizing h_1 by harmonic intervals of the 10th chord is caused by the restriction on maximal intervals considered. Indeed, under such a restriction h_1 can be recognized only by the harmonic interval $(d_1; h_1)$ which is masked by the accidental lower parallel pattern $(g; e_1)$ formed by the partials associated with g and the first overtone of e.

The failure in recognizing h_1 by melodic intervals from the previous chord, $(g; a; d_1; h_1)$, is caused by the same reason: h_1 can be detected only from melodic intervals $(h_1; h_1)$, or $(d_1; h_1)$, but the former is masked by the lower true interval $(d_1; d_1)$, while the latter—by the lower accidental pattern $(g; e_1)$.

The failure in recognizing h_1 by melodic intervals form the given chord to its successor, $(c; g; e_1; a_1)$, is also caused by masking melodic intervals $(h_1; a_1)$ and $(h_1; e_1)$, the only ones which can be used in recognition. The former is masked by the parallel accidental pattern $(d_1; c_1)$ determined by the partials associated with d_1 and the first overtone of c. The latter is masked by parallel true interval $(d_1; g)$.

Nevertheless, if we consider broader context and compare the 10th chord with the 12th chord, then the tone h_1 is recognized by melodic interval $(h_1; a_1)$ which is not masked by any pattern of a parallel interval, neither true, nor accidental.

Thus considering the hypotheses derived from confronting remote chords, one can avoid the failure in identifying notes and improve the recognition reliability.

5.7 Testing the Structural Approach

In the previous sections we have investigated the decision making approach to chord recognition. We have shown that if the chords are considered in the context and confronted to a sufficient number of other chords, they can be correctly recognized by recognizing two-tone intervals. In this section we discuss the structural approach to chord recognition which is based on simultaneously finding all the intervals of a chord, in other words, on finding the multi-interval

Figure 5.8: Chords used for testing the recognition algorithm

Table 5.10: Recognition of chord $C_7 = (c; e; g; b)$

Accuracy in semitones	Number of partials per voice			
	5	10	16	32
1	Correct	Correct	$(c; e; g; b; c_1)$	$(c; e; g; b; c_1)$
1/2	Correct	Correct	Correct	$(c; c_1; e_1; g_1; b_1; c_2; \boldsymbol{e\flat_2})$
1/3	Correct	Correct	Correct	Correct
1/4	Correct	Correct	Correct	Correct

structure of a chord.

Since recognizing intervals is the basic procedure in recognizing acoustical contours and trajectories, the conclusions concerning the efficiency and stability of recognition by simple correlation analysis are also important for the recognition by multicorrelation analysis of chord spectra. In particular, the estimates of sufficient number of patterns is important for limiting the number of the generative groups of partials (tone patterns) to be processed at the triple-correlation stage of the algorithm from Section 4.4.

The model of chord recognition described in Section 4.4 has been tested on the second series of experiments. Recall that this model is based on multicorrelation analysis of a chord spectrum and uses the criterion of least complexity of representation of spectral data in order to choose the optimal representation of the chord structure.

In our tests we have used 2 four-note chords and 3 five-note chords shown in Fig. 5.8 (we use b for *si*$\flat$ and h for *si*$\natural$). For each chord we have synthesized its Boolean spectrum, having used harmonic voices with 5, 10, 16, or 32 successive partials under the accuracy of spectral representation within 1, 1/2, 1/3, and 1/4 semitones. Having tested 5 chords in 16 different conditions, we have performed totally 80 experiments.

The recognition results for each of the five chords are displayed in Tables 5.10–5.14. Let us make some remarks concerning the recognition results referring to Table 5.13.

As seen from tracing the rows of the Table 5.13 from left to right, adding more partials to the voice pattern makes the recognition more difficult under

Table 5.11: Recognition of chord $C_{7M} = (c; e; g; h)$

Accuracy in semitones	Number of partials per voice			
	5	10	16	32
1	Correct	$(c; \boldsymbol{a\flat}; c_1; \boldsymbol{e\flat_1}; g_1)$	$(c; c_1; e_1; g_1; h_1; c_2; \boldsymbol{e\flat_2}; e_2; \boldsymbol{f_2})$	$(c; \boldsymbol{a\flat}; c_1; \boldsymbol{e\flat_1})$
1/2	Correct	Correct	$(c; e; g; h; c_1)$	$(c; \boldsymbol{a\flat}; c_1; \boldsymbol{e\flat_1})$
1/3	Correct	Correct	Correct	Correct
1/4	Correct	Correct	Correct	Correct

Table 5.12: Recognition of chord $C_{7/9} = (c; e; g; b; d_1)$

Accuracy in semitones	Number of partials per voice			
	5	10	16	32
1	Correct	$(c; e; c_1; d_1; e_1)$	$(c; e; g; b; c_1; d_1; e_1; g_1)$	$(c; e; g; b; c_1; d_1; e_1)$
1/2	Correct	$(c; e; g; d_1)$	Correct	$(c; c_1; d_1; e_1; g_1)$
1/3	Correct	$(c; e; g; d_1)$	Correct	Correct
1/4	Correct	$(c; e; g; d_1)$	Correct	Correct

Table 5.13: Recognition of chord $C_{7M/9} = (c; e; g; h; d_1)$

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

Table 5.14: Recognition of chord $C_{6/9} = (c; e; g; a; d_1)$

Accuracy in	Number of partials per voice			
semitones	5	10	16	32
1	Correct	Correct	$(c; e; g; a; c_1; d_1; e_1; g_1)$	$(c; e; g; a; c_1; d_1; e_1)$
1/2	Correct	Correct	Correct	$(c; a; c_1; d_1; e_1)$
1/3	Correct	Correct	Correct	Correct
1/4	Correct	Correct	Correct	Correct

the same accuracy of representation. In case of representation accuracy within 1 semitone, the recognition is right for the voices with 5 partials. For voices with 10 partials, only the *chord type* is recognized, i.e. the fundamental major triad $C = (c; e; g)$ instead of the full harmony $C_{7M/9} = (c; e; g; h; d_1)$. For voices with 16 partials there emerge notes which are incompatible with the original harmony. These notes are enhanced by bold face style: $\boldsymbol{a\flat}$, $\boldsymbol{b}$, and $\boldsymbol{e\flat_1}$. One can see that neither the chord, nor the chord type are correctly recognized. The harmony recognized is $A\flat_{7M/9}$ instead of original harmony $C_{7M/9}$.

As seen from tracing the columns of Table 5.13 from top to bottom, any refinement of the representation accuracy improves the recognition. In case of 32 partials per voice, the recognition is totally wrong under the representation accuracy within 1 semitone. Under the representation accuracy within 1/2 semitone, there are less notes incompatible with the original harmony. Next, under the accuracy within 1/3 semitone, the chord type is already recognizable. Finally, under the accuracy 1/4 semitone, the recognition is completely correct.

In order to compare these trends with that in human perception, we have performed audio tests with synthesized chords. These tests confirm the same trends in audio perception: The chord perceptibility is better for voices with fewer partials, and for better sharpness of hearing (we interpret the accuracy of frequency representation as the sharpness of hearing). Besides, the audio experiments show that the chords generated by tones with 32 and even 16 partials of equal power are hardly recognizable being perceived as noise. Similar difficulties in the computer recognition are seen in Tables 5.10–5.14.

Our experiments show that the chord (its contour) can be recognizable even if the generative tone spectrum is recognized with mistakes. The misrecognition of generative tones is caused by the reduction of generative patterns in our model while finding the least complex representation of spectral data (see Section 4.4). For example, translations of a generative pattern may result in coincidences of some partials from different voices, and the generative pattern can be reduced without any change in the chord spectrum (see Fig. 4.3). Such a representation is less complex than that with the original generative pattern, and therefore it is preferable. However, the pitch of such a spectral pattern can be hardly identifiable (i.e. the chord tones are not recognizable).

The recognizability of chords independently of the recognizability of tones illustrates our earlier remark, proving that high-level patterns can be recognizable even if the low-level patterns are not recognizable (see Fig. 2.2 from Section 2.2). Similarly, the type of a chord (major or minor) is recognizable regardless of the recognizability of whole chords; in our experiments we have seen that the type of chord is recognized easier than the full harmony.

Thus our model exhibits the following successive degrees in chord recognition: The chord type (major or minor) is recognized first, next goes the recognition of harmony (the chord to within permutation of notes) then the recognition of chord with its interval structure, and the chord tones are recognized worst. This precisely corresponds to human perception.

5.8 Judging Computer Experiments

Let us enumerate principal conclusions of the chapter.

1. We have realized a series of experiments on the recognition of chords with a computer model of correlative perception. At first we have investigated the basic (simple correlation) model for recognizing two-tone intervals in chords and chord progressions. Harmonic intervals of a chord are recognized by peaks of the autocorrelation function of the chord Boolean spectrum with a $\log_2$-scaled frequency axis. Melodic intervals between two chords are recognized by peaks of the correlation function of the chord Boolean spectra. The recognition of intervals is performed with no reference to absolute pitch and with no previous learning. The model is applicable to the recognition of harmonic and inharmonic voices.

2. The recognition of two-tone intervals by correlation analysis of chord spectra is tested on its efficiency and stability. It is shown that the method is more efficient and stable when the accuracy of spectral representation increases, and the intervals considered are not restricted to a certain limit. However, it is mentioned that such conditions are not realistic in practical applications. To overcome this difficulty, some special measures are proposed in order to improve the performance of the model.

3. We have proposed two approaches to improve the basic (simple correlation) model of chord recognition, decision making approach and structural approach. The final recognition by decision making approach is performed with regard to the frequency of certain tones in the totality of two-tone intervals found. The recognition by structural approach is based on finding two-tone intervals which are generated by voice patterns with the same spectral structure. It is shown that both approaches considerably improve the reliability of chord recognition.

4. The decision making approach has been tested on recognizing four-part Bach polyphony for synthesized spectra under various conditions, both for harmonic and inharmonic voices. Totally, 48 experiments have been performed. Owing to the use of the decision making option, the recognition reliability in our 48 experiments has been improved, e.g. under the accuracy of spectral representation within 1/2 semitone the recognition reliability in worst cases was improved from 93% to 98%.

5. The structural approach has been tested on recognizing four-part and five part separate chords for synthesized spectra under various conditions. Totally, 80 experiments have be performed. It has been shown that the recognition capacity of the model is close to the limits of human perception. Moreover, the trends in the performance of the model are the same as in human perception: The chord recognizability is better for voices with fewer partials and for more accurate spectral representation, interpreted as better sharpness of musical hearing. The chord type (major or minor) is recognized best, next goes the recognition of harmony (chord to within permutation of notes), then the recognition of the chord interval structure, and tones (and their pitch) are recognized worst.