# Chapter 3

# Substantiating the Model

## 3.1 On the Adequacy of the Model

In the previous chapter the principle of correlativity of perception is formulated in its general form. As already said, by correlativity of perception we understand its capability to discover similar configurations of stimuli and to form high-level configurations from them. We justify such a representation of data, supposing that in most practical cases it is less complex (requires less memory) than the totality of data.

In particular, a chord spectrum can be regarded as generated by a tone spectral pattern translated along the $\log_2$-scaled frequency axis according to some interval structure. Therefore, the main idea of our approach to chord recognition is based on finding similar groups of sinusoidal tones in the chord spectrum with which the chord can be described as an acoustical contour drawn by some generative subspectrum. We treat audio data by analogy with the visual data in Fig. 2.1 and 2.2. This analogy is displayed in Table 3.1. Note that even if the pitch is not identifiable (which corresponds to the impossibility to recognize tones like the unknown symbol II in Fig. 2.2), the chord can be recognized by relationships of the patterns of lower level (corresponding to the recognizability of symbol $B$ in Fig. 2.2).

Table 3.1: Correspondence between visual and audio data

| | In statics | | In dynamics | |
|---|---|---|---|---|
| | Visual data | Audio data | Visual data | Audio data |
| Stimuli | Pixels | Partials | Pixels | Partials |
| Low-level patterns | Symbols $A$ | Notes | Object | Notes |
| High-level pattern | Symbol $B$ | Chord | Trajectory | Melody |

Obviously, storing a tone spectrum, say, with 10 partials, and two intervals of a major triad is more efficient than storing about 30 partials of the chord spectrum. However, the efficiency of such a representation is only a conjecture. It may happen that a chord spectrum is representable in terms of generative tone spectra in several ways, implying the ambiguity in the chord recognition, or the optimal (least complex) representation may differ from the chord decomposition into notes.

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

- the description of a chord spectrum as generated by translations of a tone spectrum corresponds to the chord note structure as it is perceived;
- such a description provides the optimal representation of spectral data;
- this optimal representation is unique.

In the present chapter we prove the above items for two-tone intervals, major triads, and minor triads. This way we show that, indeed, the representation of a chord spectrum in terms of generative spectral patterns and their translations is unique, optimal, and corresponds to the note structure of the chord. In other words, we prove that our recognition model is adequate to human perception.

In Section 3.2, "Tone Spectra", we recall some basic facts from musical acoustics. We give definitions of periodical waveforms, partial tones, their amplitude and phase, and tone spectra, both discrete and continuous. Discrete spectra are considered in terms of Dirac delta-functions. Finally, the distinction between harmonic and inharmonic tones is explained.

In Section 3.3, "Representation of Tones", we enumerate the conditions for tone spectra which are assumeed in our study. We define a discrete audio power spectrum which corresponds to the one perceived by humans. For this purpose the frequency axis is logarithmically scaled and divided into a finite number of frequency bands, the phase of partial tones is ignored, and their amplitude measured by positive integers. These definitions are illustrated with examples which are used in the sequel.

In Section 3.4, "Generation of Chord Spectra", we recall the notion of convolution. We show that a chord spectrum which is generated by multiple translations of a tone spectral pattern along the $\log_2$-scaled frequency axis can be written down as a convolution product of two spectra. The first factor is a tone spectrum which is associated with a note. The second factor which is associated with the translations of the tone spectral pattern is said to be the interval distribution of the chord.

In Section 3.5, "Unique Deconvolution of Chord Spectra", we establish an isomorphism between polynomials over non-negative integers and discrete audio spectra with respect to addition and convolution. Unlike polynomials over integers, polynomials over non-negative integers have no unique factorization property, and unique deconvolution of discrete audio spectra doesn't take place in the general case. However, for spectra of two-tone intervals and major or minor triads generated by harmonic tones (with pitch salience) the unique deconvolution is valid. Hence, we obtain that the only deconvolution of a chord spectrum equals to its generation.

In Section 3.6, "Causality and Optimal Data Representation", we prove that the deconvolution of a chord spectrum into a tone spectrum and an interval distribution provides the least complex representation of spectral data. Since such a deconvolution is unique, the optimal representation of spectral data reveals the causality in their generation. From this we conclude that, in case of chord recognition, the optimal description contributes to the recognition of the causality in sound.

In Section 3.7, "Interpretation of the Results", we show that the $\log_2$-scaling of the pitch axis as well as the insensitivity of the ear to the phase of the signal are essential for adequate perception of physical reality. In particular, these properties provide for the conditions for the perception of musical tones as entire sound objects rather than as compounds of sinusoidal tones (in our model this corresponds to irreducibility of tone spectra). Besides, these assumptions are important for the capacity to recognize chords as compounds of musical tones. In other words, these conditions are necessary for recognizing the structure in audio scenes, corresponding to the physical causality in the acoustical environment.

In Section 3.8, "Main Items of the Chapter", we summarize the contents of the chapter.

## 3.2 Tone Spectra

In this section we recall some facts about tone representation. For more details one can refer to systematical introductions to musical acoustics, see e.g. Helmholtz (1877), Benade (1976), Roederer (1975), and books on signal processing, see e.g. Oppenheim & Schafer (1975); Rabiner & Gold (1975). A brief tutorial of related mathematical formulas is given by Moore (1978).

By a *musical tone*, or *harmonic tone* we understand a sound with a clear pitch salience. Such a sound is generated by a periodical, or quasi-periodical oscillation of the air pressure. Its periodical frequency, said to be the tone's *fundamental frequency* is usually associated with the tone's *pitch*.

It is known that the voices of musical instruments have a complex structure. For example, consider a cello string. Vibrating by its whole length, it produces a sinusoidal tone with fundamental frequency $\omega$. Yet the string vibrates by its halfs, thirds, etc., producing sinusoidal tones with frequencies $2\omega, 3\omega$, etc., respectively, which are said to be *partial tones* or simply *partials*, or *overtones*.

The pitches of the first 16 partials from the *overtone series* of tone $C$ (*do* of grande octave) are shown in Fig. 3.1. All these partials are summarized with different intensities, resulting in a complex oscillation perceived as the cello sound.

Note that the multiplication of the fundamental frequency implies the same difference between all the frequencies of successive partials of musical tones. This difference which is equal to the fundamental frequency $\omega$. However, the graphical distance at the stuff in Fig. 3.1 between the pitches of successive partials is getting less while increasing in their number. For example, the interval between the first two partials is an octave, whereas the interval between the second and the third partials is a fifth, etc. This is the implication of the logarithmic scale of pitch in musical notation. If a linear scale of pitch was used, then every next octave would be twice larger than the preceding one.

In signal processing the frequency axis is commonly linearly scaled. However, $\log_2$-scaled frequency axis is better suited for our purposes. Also note that such a scale is also inherent in audio perception (Gelfand 1981). It implies that equal distances on the scale correspond to equal musical intervals. The signal processing aspects of logarithmic pitch scaling are discussed by Teaney, Mourizzi, & Mintzer (1980).

According to Fourier's theorem, any tone with a periodic waveform is a sum of *harmonics*, i.e. partials with the frequencies $\omega_k$ which satisfy the *harmonic* frequency ratio

$$\omega_1 : \omega_2 : \ \dots \ : \omega_k : \ \dots = 1 : 2 : \ \dots \ : k : \ \dots . \tag{3.1}$$

Hence, any periodic waveform $f(t)$ which is a function of time $t$ can be represented as follows

$$f(t) = \sum_{k=1}^{\infty} a_k \sin(2\pi\omega_k t - \varphi_k), \tag{3.2}$$

where

$\omega_1 = \omega$ is the fundamental frequency of the waveform's oscillation which is usually associated with the tone's *pitch*;

$\omega_k = k\omega$ is the $k$th partial's frequency which is $k$ times the fundamental frequency $\omega$;

$a_k$ is the *amplitude* of the $k$th partial which is interpreted as the *loudness* of the given partial;

$\varphi_k$ is the *phase* of the $k$th partial which is interpreted as the "entry delay" of the given partial.

The formula (3.2) establishes a one-to-one correspondence between the waveform $f(t)$ and the set of characteristics of the associated sinusoidal overtones. Since the range of perceptible frequencies is approximately limited to

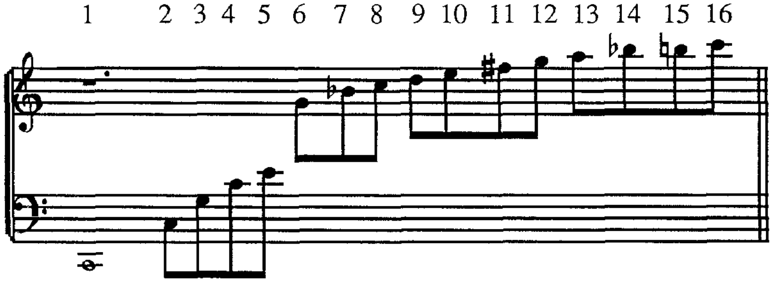



Figure 3.1: The overtone series

the band 20–20000*Hz*, the infinite series (3.2) can be always reduced to a finite sum. This means that *all* perceptible information about a musical tone is represented by a finite number of parameters.

To be more precise, the one-to-one correspondence mentioned is established between periodical functions and their spectra. The *spectrum* of a function $f(t)$ is defined to be the complex-value function

$$S(\omega) = \int_{-\infty}^{+\infty} f(t)e^{-i\omega t}\, dt.$$

If $f(t)$ is a signal, its spectrum can be understood as the density function of different sinusoidal constituents of the signal $f(t)$. Since spectral values are complex numbers, they are interpreted in terms of amplitude and phase of the related sinusoidal constituents of the signal $f(t)$.

If a function $f(t)$ satisfies some rather general conditions, then it is uniquely reconstructed from its spectrum as follows

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} S(\omega)e^{i\omega t}\, d\omega. \tag{3.3}$$

This formula justifies the interpretation of the spectrum as a density of sinusoidal constituents of the signal. In fact, by virtue of (3.3), the signal $f(t)$ is a "weighted sum" of sinusoidal partials with frequencies $\omega$ with the complex "weight coefficients" (meaning amplitude and phase) $S(\omega)$.

In case of a finite number of frequencies $\omega_1, \ldots, \omega_K$ which are present in the signal, the spectral density is concentrated at these frequencies. Then the spectrum is *pointwise*, being a sum of impulses

$$S(\omega) = \sum_{k=1}^{K} a_k \delta(\omega - \omega_k), \tag{3.4}$$

where the impulses are given by the Dirac *delta function* $\delta(x)$ conventionally defined to be

$$\delta(x-a) = \delta_a = \begin{cases} +\infty & \text{if } x = a, \\ 0 & \text{if } x \neq a, \end{cases} \tag{3.5}$$

so that

$$\int_{-\infty}^{+\infty} \delta(x-a)\,dx = 1. \tag{3.6}$$

Substituting spectrum (3.4) into (3.3), we obtain the representation (3.2). Moreover, the signal is periodical if and only if the frequency ratio of its partials is harmonic, i.e. satisfies the condition (3.1).

Usually, pointwise spectra are displayed by a series of finite (!) impulses whose height is proportional to the amplitude of related partial tones versus the frequency axis. Two such spectra are shown in Fig. 3.2b–c where the frequency axis is logarithmically scaled.

Besides harmonic tones characterized by the frequency ratio (3.1), we consider sounds with no pitch salience. These sounds can be of two types:

- *inharmonic tones*, e.g. bell-like sounds,
- *noises*, e.g. drum sounds.

Both harmonic and inharmonic *tones* are characterized by pointwise spectra. Therefore, inharmonic tones can be represented by the series (3.2) with the only difference that the ratio of partial frequencies $\omega_k$ is *inharmonic* (i.e. not harmonic).

An example of an inharmonic ratio is

$$\omega_1 : \omega_2 = 1 : \sqrt{2}.$$

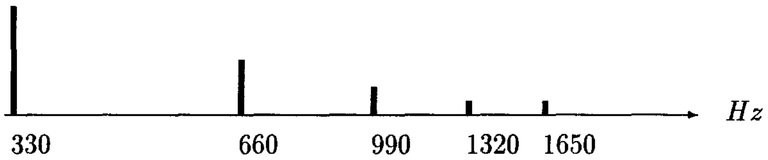
Since the above frequency ratio cannot be expressed in integers, the associated partials cannot be considered as some two partials of a harmonic tone. In particular, this implies that the signal composed by these two partials is not periodical and therefore has no salient pitch.

A noise is characterized by a *continuous* spectrum which spreads over a certain frequency band. Since all frequencies from a certain band are present, no countable set of partials is sufficient to describe the signal. Therefore, the integral representation (3.3) is needed instead of the series (3.2).

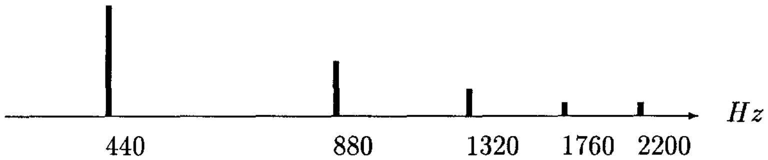
However, in discrete computational models both inharmonic tones and noises are approximated by finite sums of overtones. Therefore, in computing the distinction between inharmonic tones and noises is conditional. Recall that in discrete representations the frequency axis is divided into bands where the signal characteristics are measured. If a spectrum has constituents from successive frequency bands, such a spectrum can be considered as continuous. If a spectrum has no constituents from any pair of successive frequency bands, such a spectrum can be considered as discrete. Obviously, such a judgement depends greatly on the accuracy of discretization. Indeed, two close partials can be unseparable under a poor accuracy while being separable under a refined accuracy, implying the spectrum to be "continuous" in the former case and "discrete" in the latter.

a)

b)




c)




| | | |
|---|---|---|
| d) | $e_1$ $a_1$ $e_2$ $a_2$ $h_2$ $e_3$ $a_3$ $c\sharp_4$ | Pitch |
| e) | 0 1 2 3 4 5 6 7 8 9 10 12 14 16 18 20 22 24 26 28 30 32 34 | $n$ |
| f) | 0 8 0 0 0 0 0 0 0 0 0 0 0 4 0 0 0 0 0 0 2 0 0 0 0 1 0 0 0 1 0 0 0 0 0 | $S_n$ |
| g) | 0 0 0 0 0 0 8 0 0 0 0 0 0 0 0 0 0 0 0 4 0 0 0 0 0 0 2 0 0 0 0 1 0 0 0 1 | $[\delta_5 * S]_n$ |

Figure 3.2: Example of tone spectra

*a)* tones $e_1$ and $a_1$ in standard musical notation;

*b)* the audio spectrum (with $\log_2$-scaled frequency axis) of tone $e_1$ for a harmonic voice with 5 successive partials which have decreasing power;

*c)* the same for tone $a_1$;

*d)* the pitches $e_1, \ldots, c\sharp_4$ which correspond to the mean frequencies of the frequency bands for the given spectral representations;

*e)* the indexes $n = 1, \ldots, 34$ of frequency bands for the given discrete spectra;

*f)* the discrete audio spectrum of tone $e_1$ under the frequency resolution within a semitone (1/12 octave) — string $S_n$;

*g)* the same for the second tone — string $[\delta_5 * S]_n$.

Since we deal with discrete models, we make the only principal distinction between harmonic tones, characterized by the harmonic ratio of partial frequencies (3.1), and all other sounds which are said to be *inharmonic sounds.* Moreover, this distinction is true within certain accuracy.

## 3.3 Representation of Tones

Thus in the present study we consider *audio* spectra, i.e. we assume that the frequency axis is $\log_2$-scaled, implying equal distances corresponding to equal musical intervals.

We restrict ourselves to the spectra which are *limited in low frequencies* by a fixed threshold, say, by $20Hz$ and *bounded in high frequencies* by variable thresholds, depending on the spectrum.

Since man is not sensitive to the phase of the signal (Gelfand 1981), we restrict our consideration to *power spectra* where the phase of the overtones is ignored (and well as the phase $180^0$ which corresponds to negative values of partials' amplitude).

Finally, the spectra to be considered are *discrete*, i.e. it is assumed that the frequency range is divided into bands wherein the signal amplitude is measured by positive integers.

All of this means that both frequency bands and spectral values can be enumerated by non-negative integers, while the number of bands with a strictly positive level being always finite.

Thus by *spectra* we understand the expressions of the form

$$S = S(x) = \sum_{n=0}^{N} S_n \delta(x - n) = \sum_{n=0}^{N} S_n \delta_n, \tag{3.7}$$

where

$N$ is the total number of frequency bands minus 1;

$n$ is the index of a frequency band;

$\delta_n$ is the Dirac delta function defined in (3.5–3.6), i.e. the unit impulse at the $n$th frequency;

$S_n$ is a non-negative integer interpreted as the signal power in the $n$th frequency band.

The *support* of spectrum (3.7) is defined to be the set of its partial frequencies. To be more precise, it is defined as the set of indexes of frequency bands as follows

$$\Delta_S = \{n : S_n \neq 0\}.$$

Since by assumptions we consider bounded spectra, the spectral supports are always finite.

By virtue of (3.1), the $k$th partial of a harmonic tone falls into the frequency band with the following index

$$n_k = p + [C \log_2 k + 0.5], \tag{3.8}$$

where

$p = n_1$ is the index of the frequency band with the tone's *fundamental frequency*;

$C$ is the constant which characterizes the accuracy of spectral representation, being equal to the number of frequency bands per octave,

$[\cdot + 0.5]$ is the *rounding function* (since $[\cdot]$ retains the integer part of its argument, function $[\cdot + 0.5]$ retains the closest integer to a given real number).

A spectrum with the partials which satisfy the condition (3.8) is said to be *harmonic*. It is easy to see that a discrete audio spectrum $S$ is harmonic if and only if its support $\Delta_S$ has only harmonic frequencies, i.e. for a certain $K$ it holds

$$\Delta_S \subset \{n_k : n_k = [C \log_2 k + 0.5],\ k = 1, \ldots, K\}. \tag{3.9}$$

Two harmonic spectra are shown in Fig. 3.2.

## 3.4 Generation of Chord Spectra

First of all, recall that the *convolution* of two functions $f(t)$ and $g(x)$ is defined to be the function

$$[f * g](t) = \int_{-\infty}^{+\infty} f(t - x) g(x)\, dx.$$

Under rather general conditions the convolution is commutative, associative, and linear with respect to both arguments $f$ and $g$.

As follows from the definition of delta function (3.5–3.6),

$$[f * \delta_a](t) = \int_{-\infty}^{+\infty} f(t - x)\delta(x - a)\, dx = f(t - a),$$

corresponding to the translation of function $f$ by $a$ to the right along the $t$-axis.

Consequently, the delta function $\delta = \delta_0$ is a *convolution unit*, since for all functions $f$ it holds

$$\delta * f = f * \delta = f.$$

Also note that for arbitrary $a$ and $b$ we have

$$\delta_a * \delta_b = \delta_{a+b}.$$

Since we consider a $\log_2$-*scaled* frequency axis, a pitch shift of a complex tone spectrum corresponds to a *parallel* translation of the whole tone spectrum along the frequency axis. A translation of a spectrum (3.7) by $m$ bands to the right is given by the convolution

$$\delta_m * S = \sum_n S_n \delta_{n+m} \tag{3.10}$$

An example of a translation of a harmonic spectrum is shown in Fig. 3.2.

As fixed by Conjecture 4 from Section 1.4, a chord spectrum can be regarded as generated by a multiple translation of a tone spectrum. By virtue of (3.10), a chord spectrum $S$ can be written down as follows

$$S = \sum_m [I_m \delta_m * T] = [\sum_m I_m \delta_m] * T = I * T, \tag{3.11}$$

where $T$ is a tone spectrum, and

$$I = \sum_m I_m \delta_m$$

is said to be the *interval distribution* of the chord.

For example, the interval distribution associated with the interval of $m$ semitones is defined to be

$$I_m = \delta_0 + \delta_{[mC/12+0.5]}.$$

If the accuracy of frequency resolution is the same as the accuracy of standard note stuff, i.e. to within one semitone ($C = 12$), then the major triad (4 semitones) has the interval distribution

$$I_4 = \delta_0 + \delta_4. \tag{3.12}$$

Now consider a chord which is formed by two intervals of $m$ and $k$ semitones from the lowest tone. The corresponding interval distribution is as follows

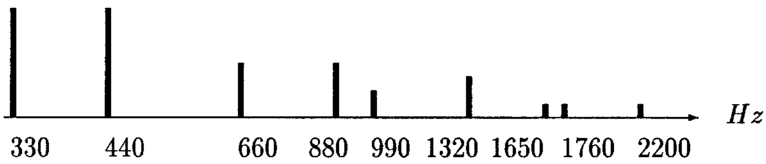
$$I_{m,k} = \delta_0 + \delta_{[mC/12+0.5]} + \delta_{[kC/12+0.5]}.$$

If the accuracy of frequency resolution is to within a quarter of a tone ($C = 24$), then the interval distribution of the major triad which is built by major third and fifth (4 and 7 semitones) has the form

$$I_{4,7} = \delta_0 + \delta_8 + \delta_{14}.$$

*a*)

*b*)




| | | |
|---|---|---|
| *c*) | $e_1$ $a_1$ $e_2$ $a_2$ $h_2$ $e_3$ $a_3$ $c\sharp_4$ | Pitch |
| *d*) | 0 1 2 3 4 5 6 7 8 9 10 12 14 16 18 20 22 24 26 28 30 32 34 | $n$ |
| *e*) | 0 8 0 0 0 0 8 0 0 0 0 0 0 4 0 0 0 0 4 0 2 0 0 0 0 3 0 0 0 1 1 0 0 0 1 | $S_n$ |

Figure 3.3: Example of chord spectrum

*a*) chord $(e_1, a_1)$ in standard musical notation;

*b*) the audio spectrum (with $\log_2$-scaled frequency axis) of the chord for a harmonic voices with 5 successive partials which have decreasing power;

*c*) the pitches $e_1, \ldots, c\sharp_4$ which correspond to the mean frequencies of the frequency bands for the given spectral representations;

*d*) the indexes $n = 1, \ldots, 34$ of frequency bands for the given discrete spectra;

*e*) the discrete audio spectrum of the chord under the frequency resolution within a semitone (1/12 octave) — string $S_n$.

A spectrum $S$ is said to be *simple* if all its coefficients are relatively prime (in the sense of number theory) and its first coefficient $S_0 \neq 0$.

Obviously, the decomposition (3.11) can be written down as follows

$$S = a\delta_p * T * I, \tag{3.13}$$

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 $I_m$ determining their *relative loudness*. The term $a\delta_p * T$ can be understood as a spectrum of the lowest tone of the chord with spectral pattern $T$, loudness $a$, and pitch $p$ (the index of the frequency band with the tone's fundamental frequency).

For example, in Fig. 3.3 one can see a chord spectrum which is composed of two tone spectra shown in Fig. 3.2. The spectrum $S$ of the chord from Fig. 3.3 can be written down in the form (3.13) as follows

$$S = \delta_1 * (8\delta_0 + 4\delta_{12} + 2\delta_{19} + \delta_{24} + \delta_{28}) * (\delta_0 + \delta_5).$$

## 3.5 Unique Deconvolution of Chord Spectra

By *deconvolution* of a spectrum $S$ we shall understand a representation of $S$ as a convolution product of two or more spectra

$$S = S_1 * \ldots * S_k,$$

where no factor $S_i$ is a convolution unit, i.e. $\delta$ multiplied by a constant $a$.

In order to investigate the items enumerated in section 3.1, we pose the question: Given a chord spectrum $S$ generated according to (3.13); does there exist a deconvolution of $S$ into simple spectra other than (3.13)?

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

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

*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.*

PROOF OF LEMMA 1. Obviously, the correspondence is one-to-one. Consider two spectra and the associated polynomials

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

$$T = \sum_{n=0}^{M} b_n \delta_n \quad \longleftrightarrow \quad q(x) = \sum_{n=0}^{M} b_n x^n.$$

Then

$$S + T = \sum_{n=0}^{\max\{M,N\}} (a_n + b_n)\delta_n \longleftrightarrow \sum_{n=0}^{\max\{M,N\}} (a_n + b_n)x^n = p(x) + q(x).$$

Since

$$x^{i+j} = x^i x^j \longleftrightarrow \delta_i * \delta_j = \delta_{i+j},$$

all the difference between multiplication of polynomials and convolution of spectra is in manipulating lower indexes of $\delta$ instead of powers of $x$. Hence, we have

$$S * T = \sum_{n=0}^{M+N} \Big( \sum_{i+j=n} a_i b_j \Big)\delta_n \longleftrightarrow \sum_{n=0}^{M+N} \Big( \sum_{i+j=n} a_i b_j \Big)x^n = p(x)\, q(x),$$

as required. ■

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 convolution unit $\delta$ multiplied by a constant $a$.

Referring to the unique factorization of polynomials over integers (Birkhoff & Mac Lane 1965), we obtain the unique deconvolution of spectra with integral coefficients. Since we are interested in the deconvolution of discrete spectra with *non-negative* integral coefficients into spectra with *non-negative* integral coefficients, the question of unique deconvolution is reformulated for polynomials with the same property.

Ambiguous factorization into primes arises in special number fields, see Pollard (1950), p. 76, and, as shown at the end of this section, ambiguous factorization is also inherent in polynomials over non-negative integers. However, the polynomials associated with discrete power spectra of usual chords, like two-tone intervals and major or minor triads, are uniquely factored into irreducible polynomials. In order to prove the unique deconvolution theorem for chord spectra, we need several propositions.

**Lemma 2 (Sufficient Condition for Irreducible Spectra)** *Consider a simple spectrum $S$ whose support $\Delta_S$ contains two points at least, i.e.*

$$S = \sum_{k=1}^{K} a_k \delta_{\omega_k}, \quad \omega_1 < \ldots < \omega_K, \ K \geq 2.$$

*Suppose that the distance between the last two partials (first two partials) of $S$ is less than the distance between any other pair of partials of $S$, i.e.*

$$d = \omega_K - \omega_{K-1} < \omega_k - \omega_{k-1}, \ k = 2, \ldots, K-1 \tag{3.14}$$

$$(d = \omega_2 - \omega_1 < \omega_k - \omega_{k-1}, \ k = 3, \ldots, K). \tag{3.15}$$

*Then $S$ is irreducible.*

PROOF OF LEMMA 2. Firstly, assume that the distance $d$ between the last two partials in $S$ is less then the distance between any other pair of partials. Since $S$ is simple, the factorization $S = \delta_p * S'$, where $p > 0$, is impossible. Consequently, if $S = I * T$ then each of the supports $\Delta_I$ and $\Delta_T$ should contain two points at least. Since the partials indexed by $\omega_{K-1}$ and $\omega_K$ are the last in the spectrum $S$, and operator $I * (\cdot)$ performs parallel translations of spectrum $T$, only the following two cases are possible.

1. The partials indexed by $\omega_{K-1}$ and $\omega_K$ result from the two last partials of spectrum $T$ translated. Then the distance between the last two partials of $T$ is also equal to $d$. Since by assumption $\Delta_I$ contains more than one point, the spectrum $T$ is repeated in $S$ several times. Consequently the same distance between partials $d$ appears elsewhere in $S$, which contradicts the condition (3.14).

2. The partials indexed by $\omega_{K-1}$ and $\omega_K$ result from two translations of the last partial of spectrum $T$. Then the distance between the last two points in $\Delta_I$ is equal to $\omega_K - \omega_{K-1}$. Since by assumption $\Delta_T$ contains more than one point, the same distance between points should appear elsewhere in $\Delta_S$, resulting from the same translation of the first partial of $T$. This however contradicts the condition (3.14).

Thus we have shown the impossibility of convolution factoring $S$ into two non-trivial spectra. One can easily modify the proof for the assumption (3.15). ■

**Lemma 3 (Irreducibility of Harmonic Spectra)** *Let $T$ be a simple harmonic spectrum or a simple segment of a harmonic spectrum, i.e.*

$$T = \sum_{k=K_1}^{K} a_k \delta_{\log_2 k}, \quad 1 \leq K_1 < K.$$

*Then $T$ is irreducible.*

PROOF OF LEMMA 3. Consider the support of spectrum $T$. Note that the difference $\log_2 k - \log_2(k-1)$ decreases monotonically with an increase in $k$. Consequently, the distance between the last two partials of $T$ is less than the distance between any other pair of partials. Then by Lemma 2 we obtain the required statement. ■

Note that in case of discrete spectra the above lemma is valid, if the spectral resolution is sufficiently accurate to distinguish that the distance between the last two partials is less than the distance between any other pair of partials.

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

PROOF OF LEMMA 4. The required fact for intervals follows from Lemma 2. Since the major and minor triads are formed by major and minor thirds, the corresponding interval distributions satisfy the assumption of Lemma 2. ■

Obviously, the above lemma is valid for discrete spectra as well.

Also note that the assumption of non-negativity of spectral coefficients is important. Otherwise, as seen from the following example, even the interval of major triad is not irreducible.

**Example 1 (Reducibility of Major Triad Interval Distribution)** Let the spectral accuracy be within one semitone ($C = 12$). Hence, the major triad corresponds to 4 frequency bands. Then by virtue of (3.12) and Lemma 1, the reducibility of the major triad follows from the polynomial factorization

$$x^4 + 4 = (x^2 + 2x + 2)(x^2 - 2x + 2). \tag{3.16}$$

Further we shall divide a spectrum

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

whose partial (impulses) 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:

$$S = \sum_{n=1}^{N} a_n \delta_{\omega_n} = \sum_{n=1}^{N-Q} a_n \delta_{\omega_n} + \sum_{n=N-Q+1}^{N} a_n \delta_{\omega_n},$$

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

$$S_Q = \sum_{n=N-Q+1}^{N} a_n \delta_{\omega_n}$$

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

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

$$S_Q = \delta_a * T_Q$$

for certain $a$. 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*

$$T * I = U * J.$$

*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*

$$T_Q = \sum_{n=N-Q+1}^{N} a_n \delta_{\omega_n} = \sum_{\omega_n : \omega_n > \omega_N - f} a_n \delta_{\omega_n}, \tag{3.17}$$

*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 $U_Q \sim T_Q$ and the distance between the last two partials of $J$ is greater than or equal to $g$, or $J_Q \sim T_Q$ and the distance between the last two partials of $U$ is greater than or equal to $g$.*

Proof of Lemma 5. Consider spectrum

$$S = T * I.$$

Since $S$ is generated by translations of $T$ at intervals from $I$, and the distance $f$ between the last two impulses of $I$ determines the size of $T_Q$, we have

$$S_Q \sim T_Q. \tag{3.18}$$

In particular, the distance between the last two impulses of $S_Q$ is equal to $d$ which is the smallest distance between impulses of $T_Q$.

Consider the last two impulses of $S$. Since by assumption $f > d$, we have

$$S_2 \sim T_2.$$

By assumption $U$ and $J$ have at least two impulses each. Let us analyze several possibilities.

The case when the distance between the last two impulses of $U$ is strictly smaller than $d$ is impossible, since then the distance between the last two impulses of $S = U * J$ would be less than $d$.

By the same reasons, the distance between the last two impulses of $J$ cannot be strictly smaller than $d$.

The case when both the distance between the last two impulses of $U$ and the distance between the last two impulses of $J$ are strictly greater than $d$ is also impossible, since then the distance between the last two impulses of $S = U * J$ would be strictly greater than $d$.

Therefore, the only possible cases are the following:

(a) The distance between the last two impulses of $U$ is equal to $d$ and the distance between the last two impulses of $J$ is strictly greater than $d$.

(b) The distance between the last two impulses of $J$ is equal to $d$ and the distance between the last two impulses of $U$ is strictly greater than $d$.

Since the above two cases differ in renaming $U$ and $J$, consider the first case. Let the partials (impulses) of $S$ be indexed from the end of $S$ by numbers $q = 1, \dots, Q$ (the partial indexed by $q$ is the $q$th to last partial of $S$). Now we shall show that every $q$th to last impulse of $S$ results from the convolution of the $q$th to last impulse of $U$ by the last impulse of $J$. We shall prove it by induction on $q = 2, \dots, Q$.

For $q = 2$ the required statement is already proved. Suppose that it is valid for a certain positive integer $q < Q$. Obviously, for the $q + 1$st to last impulse of $S$ there are only four possibilities:

1. The $q+1$st to last impulse of $S$ results from the convolution of the $q+1$st to last impulse of $U$ by the last impulse of $J$.
2. The $q+1$st to last impulse of $S$ results from the convolution of the $q+1$st to last impulse of $U$ by a not-last impulse of $J$.
3. The $q + 1$st to last impulse of $S$ results from the convolution of the last impulse of $J$ by an impulse of $U$ whose number from the end is $r < q+1$.
4. The $q+1$st to last impulse of $S$ results from the convolution of a not-last impulse of $J$ by an impulse of $U$ whose number from the end is $r < q+1$.

Consider the second case. If the $q + 1$st to last impulse of $S$ results from the convolution of the $q + 1$st impulse of $U$ by a not-last impulse of $J$, then, since $S = U * J$, the distance $d$ must be repeated in $S_Q$, which contradicts to the fact that by virtue of (3.18) the distance $d$ is the smallest distance between partials of $S_Q$. Consequently, this case is impossible.

Consider the third case. By the statement of induction the convolution of the $r$th to last impulse of $U$ by the last impulse of $J$ is equal to the $r$th to last impulse of $S$. Consequently, the third case is impossible.

The fourth case is impossible by the same reasons as the second case.

Therefore, the only possible case is the first, whence $U_{q+1} \sim S_{q+1}$.

Thus we have shown that $U_Q \sim S_Q$, whence by virtue of (3.18) we have

$$U_Q \sim T_Q.$$

Since every impulse of $S_Q$ results from the convolution of the last impulse of $J$ by the corresponding impulse of $U$, the distance between the last two impulses of $J$ is greater than or equal to $g$ (otherwise the shortest distance $d$ would be repeated in $S_Q \sim T_Q$). ■

**Lemma 6 (Uniqueness of Interval Decomposition)** *Consider an interval distribution $I$ with two impulses, and let the distance $f$ between its impulses be smaller than or equal to 12 semitones, i.e.*

$$\begin{aligned} I &= a_1\delta_{\omega_1} + a_2\delta_{\omega_2} \quad (\omega_1 < \omega_2); \\ f &= \omega_2 - \omega_1 \leq 12 \textit{ semitones.} \end{aligned}$$

*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.*

$$\begin{aligned} T &= \sum_{n=1}^{N} b_n\delta_{\omega_n} \quad (\omega_1 < \ldots < \omega_N); \\ f &> \omega_N - \omega_{N-2}. \end{aligned}$$

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

Proof of Lemma 6. Consider a deconvolution of $S$ into two non-trivial factors $U$ and $J$, i.e. $U * J = S$. By virtue of Lemma 5 we can assume that

$$U_3 \sim T_3.$$

We are going to show that $J = I$ whence it will follow that $U = T$.

At first let us show that $J$ has two impulses as $I$ has, and that the distance between these impulses is also $f$. Suppose that it is not true. Since by assumption $J$ has two impulses at least, there exists an impulse in $J$ whose distance from the highest impulse of $J$ differs from $f$, and therefore the tail $T_3$ is repeated in $S$ at a distance from the end of $S$ different from $f$. We shall show that it is not possible.

Consider spectrum $S$ as a union of two sets of harmonics. Let the first set correspond to the lowest tone with the fundamental frequency $p$. Then the frequencies of its harmonics are $np$, where $n = 1, \ldots, N$. Let the second set correspond to the upper tone with the fundamental frequency $q$. Then the frequencies of its harmonics are $kq$, where $n = 1, \ldots, N$. Since by assumption the interval considered is not greater than the octave, we have

$$p < q \leq 2p. \tag{3.19}$$

Since we suppose that $T_3$ is repeated in $S$ at an interval different from $f$, there exists a positive frequency

$$r \neq p, q \tag{3.20}$$

such that the frequencies $Nr$, $(N-1)r$, and $(N-2)r$ are equal to some frequencies of spectrum $S$. Since $Nq$ is the highest frequency of spectrum $S$, we have $Nr \leq Nq$, whence by virtue of (3.20)

$$r < q. \tag{3.21}$$

Since we suppose that frequency $Nr$ is inherent in spectrum $S$, there are two possibilities:

1. $Nr$ is a harmonic frequency of the higher tone, i.e.

$$Nr = kq, \ k < N, \tag{3.22}$$

(if $k = N$ we would have $r = q$ against (3.20)).

2. $Nr$ is a harmonic frequency of the lower tone, i.e.

$$Nr = mp, \ m < N, \tag{3.23}$$

(if $m = N$ we would have $r = p$ against (3.20)).

Let us analyze both possibilities.

1. Consider possibility (3.22). Since both frequencies $(N-1)r$ and $(N-2)r$ are inherent in spectrum $S$, we have the following cases.

   (a) $(N-1)r = k_1 q$, where obviously $k_1 < k$.
   By virtue of (3.22) this gives

   $$r = (k - k_1)q, \quad k - k_1 \geq 1,$$

   whence $r \geq q$ against (3.21). Therefore, this case is impossible.

   (b) Consequently, it must be $(N-1)r = m_1 p$ for a certain $m_1$.

   (c) $(N-2)r = m_2 p$, where obviously $m_2 < m_1$.
   By virtue of (b), this gives

   $$r = (m_1 - m_2)p, \quad m_1 - m_2 \geq 1,$$

   whence either $r = p$ against (3.20), or $r \geq 2p$ against (3.19) and (3.21).

   (d) $(N-2)r = k_2 q$, where obviously $k_2 < k$.
   By virtue of (3.22) this gives

   $$r = \frac{k - k_2}{2} q, \quad k - k_2 \geq 1.$$

The case $\frac{k-k_2}{2} \geq 1$ contradicts to (3.21), consequently $\frac{k-k_2}{2} = \frac{1}{2}$, which is equivalent to $k_2 = k - 1$. Substituting it into (c) and using (3.22), we obtain

$$\frac{N}{N-2} = \frac{k}{k-1}, \quad k - k_2 \geq 1,$$

which is equivalent to

$$1 + \frac{2}{N-2} = 1 + \frac{1}{k-1},$$

whence

$$N = 2k.$$

Since by assumption $N$ is odd, this is impossible.

Therefore, the possibility $Nr = kq$ is excluded.

2. Consider possibility (3.23). Since both frequencies $(N-1)r$ and $(N-2)r$ are inherent in spectrum $S$, we have the following cases.

   (e) $(N-1)r = m_1 p$, where obviously $m_1 < m$.
   By virtue of (3.23) this gives

   $$r = (m - m_1)p, \quad m - m_1 \geq 1,$$

   whence either $r = p$ against (3.20), or $r \geq 2p$ against (3.19) and (3.21). Therefore, this case is impossible.

   (f) Consequently, it must be $(N-1)r = k_1 q$ for a certain $k_1$.

   (g) $(N-2)r = k_2 q$, where obviously $k_2 < k_1$.
   By virtue of (f) this gives

   $$r = (k_1 - k_2)q, \quad k_1 - k_2 \geq 1,$$

   whence $r \geq q$ against (3.21).

   (h) $(N-2)r = m_2 p$, where obviously $m_2 < m$.
   From this and (3.23) we obtain

   $$r = \frac{m - m_2}{2}p, \quad m - m_2 \geq 1.$$

   The case $\frac{m-m_2}{2} \geq 2$ implies $r \geq 2p \geq q$ against (3.21).
   The case $\frac{m-m_2}{2} = 1$ implies $r = p$ against (3.20).
   The case $\frac{m-m_2}{2} = \frac{1}{2}$ implies $m_2 = m - 1$. Substituting it into (h) and using (3.23), we obtain

   $$\frac{N}{N-2} = \frac{m}{m-1},$$

which is equivalent to

$$N = 2m,$$

which is impossible since by assumption $N$ is odd.

Finally, the case $\frac{m-m_2}{2} = \frac{3}{2}$ implies $m_2 = m - 3$. Substituting it into (h) and using (3.23), we obtain

$$\frac{N}{N-2} = \frac{m}{m-3},$$

which is equivalent to

$$1 + \frac{2}{N-2} = 1 + \frac{3}{m-3},$$

whence

$$3N = 2m$$

which is also impossible since by assumption $N$ is odd.

Therefore, the possibility $Nr = mp$ is excluded as well.

Thus a 3-tail congruent to $T_3$ can appear in $S$ only at the distance $f$ from the last partial of $S$.

Now let us show that $J = I$ to within a convolution unit. Consider the tail $T_2$ with the distance $d$ between its two impulses. Since $S$ contains two subspectra congruent to $T_2$, namely $\delta_a * T_2$ (for certain $a$) and $\delta_{a+f} * T_2$, and the distance between the $T_2$'s impulses $d < f$, one of impulses of $T_2 * \delta_a$ is not superimposed on any harmonic of $T * \delta_{a+f}$. This not-superimposed impulse of $T_2 * \delta_a$ can be confronted to the corresponding impulse of $T * \delta_{a+f}$ which by Lemma 5 is also superimposed on no other impulse of lower tones. Consequently, the ratio of these "pure" harmonics determines the ratio of coefficients in $I$.

Thus $J = I$ to within units. By virtue of Lemma 1 and uniqueness of polynomial division, this implies $U = T$ to within a convolution unit. By Lemmas 3 and 4 factors $I$ and $T$ are irreducible and no further deconvolution of $J$ and $U$ is possible, as required. ■

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.*

$$\begin{aligned} I &= a_1\delta_{\omega_1} + a_2\delta_{\omega_2} \quad (\omega_1 < \omega_2); \\ f &= \omega_2 - \omega_1 \leq 12 \textit{ semitones}. \end{aligned}$$

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

$$\begin{aligned} T &= \sum_{n=1}^{N} b_n \delta_{\omega_n} \quad (\omega_1 < \ldots < \omega_N); \\ f &> \omega_N - \omega_{N-2}. \end{aligned}$$

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

PROOF OF THEOREM 1. By virtue of Lemma 6 it suffices to consider the case when $N$ is even. Consider a deconvolution of $S$ into two non-trivial factors $U$ and $J$, i.e. $U * J = S$. By virtue of Lemma 5 we can assume that

$$U_4 \sim T_4.$$

Now consider $U_4$ without its last partial, and repeat the proof of Lemma 6 for three frequencies $(N-1)r$, $(N-2)r$, and $(N-3)r$, substituting $N' = N - 1$ for $N$ everywhere in the proof of Lemma 6. Since by assumption $N$ is even, $N'$ is odd, and the proof remains valid. ■

**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.*

$$\begin{aligned} I &= \sum_{i=1}^{3} a_i \delta_{\omega_i} \quad (\omega_1 < \omega_2 < \omega_3); \\ f &= \omega_3 - \omega_1 \leq 12 \textit{ semitones}, \\ \omega_2 - \omega_1 &\neq \omega_3 - \omega_2. \end{aligned}$$

*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 four harmonics of $T$ is less than $f$, i.e.*

$$\begin{aligned} T &= \sum_{n=1}^{N} b_n \delta_{\omega_n} \quad (\omega_1 < \ldots < \omega_N); \\ f &> \omega_N - \omega_{N-3}. \end{aligned}$$

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

PROOF OF LEMMA 7. The proof follows the ideas of Lemma 6. The difference is that instead of two possibilities (3.22) and (3.23) one must consider the third possibility, $Nr = ls$, where $s$ is the fundamental frequency of the middle tone of the chord, i.e. $p < s < q$, and $l$ is a positive integer such that $l < N$. Moreover, one has to add some more cases to (a)–(d) and (e)–(h), because the frequencies $nr$ can fall onto the harmonics of this third tone with fundamental frequency $s$.

Suppose that in $S$ there exists a 4-tail congruent to $T_4$ which is different from the 4-tails of the chord tones. Denote its fundamental frequency by $r$, $r \neq p, q, s$, and consider four partials $Nr$, $(N-1)r$, $(N-2)r$, $(N-3)r$, which by our hypothesis are inherent in spectrum $S$. Since these four partials are inherent in chord spectrum $S$ built from harmonic tones with fundamental frequencies $p$, $q$, and $s$, at least two of these four partials fall on harmonics of the same tone.

The case when two successive partials, say $Nr$ and $(N-1)r$, are harmonic frequencies of the same tone of the chord is brought to contradiction in the same way as in the proof of Lemma 6.

The case when a partial and the next to next partial, say $Nr$ and $(N-2)r$, are harmonic frequencies of the same tone of the chord is also brought to contradiction in the same way as in the proof of Lemma 6.

The only new case arises when $Nr$ and $(N-3)r$ are harmonic frequencies of the same tone, but $(N-1)r$ and $(N-2)r$ are harmonic frequencies of two other tones of the chord.

At first suppose that $Nr$ falls onto a harmonic of the upper tone of the chord, i.e. $Nr = kq$ and $(N-3)r = k_3q$, where $N > k > k_3$. Using the reasons similar to that from the proof of Lemma 6, we obtain

$$r = \frac{k-k_3}{3}q, \quad k - k_3 \geq 1.$$

The case $k - k_3 \geq 3$ implies $r \geq q$ against (3.21).

The case $k - k_3 = 1$ implies

$$\frac{N}{N-3} = \frac{k}{k-1},$$

which is equivalent to

$$N = 3k,$$

which is impossible since by assumption $N$ is not divisible by 3.

The case $k - k_3 = 2$ implies

$$\frac{N}{N-3} = \frac{k}{k-2},$$

which is equivalent to

$$2N = 3k,$$

which is also impossible since by assumption $N$ is not divisible by 3.

Now suppose that $Nr$ falls onto a harmonic of the lower or middle tone of the chord. For example, let $Nr = ls$ and $(N-3)r = l_3 s$, where $N > l > l_3$. This implies

$$r = \frac{l - l_3}{3}s, \quad l - l_3 \geq 1.$$

The case $l - l_3 \geq 6$ implies $r \geq 2s \geq 2p \geq q$ against (3.19) and (3.21).

The case $l - l_3 = 3$ implies $r = s$ against $r \neq s$.

The case $l - l_3 = 1$ implies

$$\frac{N}{N-3} = \frac{l}{l-1},$$

which is equivalent to

$$N = 3l,$$

which is impossible since by assumption $N$ is not divisible by 3.

The case $l - l_3 = 2$ implies

$$\frac{N}{N-3} = \frac{l}{l-2},$$

which is equivalent to

$$2N = 3l,$$

which is also impossible since by assumption $N$ is not divisible by 3.

The case $l - l_3 = 4$ implies

$$\frac{N}{N-3} = \frac{l}{l-4},$$

which is equivalent to

$$4N = 3l,$$

which is impossible since by assumption $N$ is not divisible by 3.

The case $l - l_3 = 5$ implies

$$\frac{N}{N-3} = \frac{l}{l-5},$$

which is equivalent to

$$5N = 3l,$$

which is also impossible since by assumption $N$ is not divisible by 3.

Thus in $S$ any 4-tail congruent to $T_4$ falls onto 4-tails of the chord tones. The remainder of the proof is similar to that of Lemma 6. ∎

**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.*

$$\begin{aligned} I &= \sum_{i=1}^{3} a_i \delta_{\omega_i} \quad (\omega_1 < \omega_2 < \omega_3); \\ f &= \omega_3 - \omega_1 \leq 12 \textit{ semitones}, \\ \omega_2 - \omega_1 &\neq \omega_3 - \omega_2. \end{aligned}$$

*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.*

$$\begin{aligned} T &= \sum_{n=1}^{N} b_n \delta_{\omega_n} \quad (\omega_1 < \ldots < \omega_N); \\ f &> \omega_N - \omega_{N-3}. \end{aligned}$$

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

PROOF OF THEOREM 2. The statement of the theorem follows from Lemma 7 in the same way as the statement of Theorem 1 follows from Lemma 6. For $N$, we substitute the greatest $N' \leq N$ which is not divisible by 2 and by 3. Since obviously $N' \geq N - 3$, and the above lemma is formulated for 4-tails, the seven partials in the tail $T_7$ is sufficient to adapt the lemma in order to prove the theorem. ■

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 deconvolution of a chord spectrum, which is built from harmonic tones with a sufficient number of harmonics, reveals its generation.

Further generalizations of the Unique Decomposition Theorem to multi-note chords can be formulated by analogy with the generalization of Lemma 6 and Theorem 1 to Lemma 7 and Theorem 2, respectively.

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 Deconvolution Theorem. For arbitrary power spectra the unique deconvolution doesn't hold. By virtue of Lemma 1 this is seen from the following example proposed by Chateauneuf (1993) in personal communication.

**Example 2 (No Unique Factorization of Polynomials over Positive Integers)** Consider the following polynomials:

$$\begin{aligned} p(x) &= x^2 + 2x + 2 \\ q(x) &= x^2 - 2x + 2 \\ r(x) &= x(x^2 + 2x + 2) + 1 = (x^2 + x + 1)(x + 1), \end{aligned}$$

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. According to the reasons which precede (3.16), polynomial

$$p(x)q(x) = x^4 + 4$$

is irreducible over non-negative integers. Therefore, polynomial with non-negative integer coefficients $p(x)q(x)r(x)$ can be factored into irreducible polynomials over non-negative integers as follows

$$p(x)q(x)r(x) = [p(x)q(x)]r(x) = (x^4 + 4)(x^2 + x + 1)(x + 1). \tag{3.24}$$

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

$$\begin{aligned} p(x)q(x)r(x) = p(x)[q(x)r(x)] &= (x^2 + 2x + 2)[x(x^4 + 4) + x^2 - 2x + 2] \\ &= (x^2 + 2x + 2)(x^5 + x^2 + 2x + 2), \end{aligned} \tag{3.25}$$

where the first factor, $x^2 + 2x + 2$, is irreducible, being different from all irreducible factors of factorization (3.24). Hence, there exist two different factorizations of polynomial $p(x)q(x)r(x)$ into irreducible polynomials over non-negative integers, one given by (3.24) and another given by (3.25) with a further factorization of the second term $x^5 + x^2 + 2x + 2$, if such a further factorization exists.

## 3.6 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 (Kolmogorov 1965; Calude 1988). 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 $\Delta_S$. The complexity of $S$ is denoted by $|\Delta_S|$.

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

**Theorem 3 (Revealing Causality by Optimal Data Representation)** *Suppose that a spectrum $S$ is generated by a spectrum $T$ translated according to an interval 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 (3.13) is the least complex representation of $S$.*

PROOF OF THEOREM 3. By Theorem 2 there is a unique deconvolution of a chord spectrum $S$ given by its generation (3.13). Therefore, the problem is brought to comparing the complexity $|\Delta_S|$ of the spectrum $S$ and the complexity of its deconvolution which is

$$|\Delta_{a\delta_p}| + |\Delta_T| + |\Delta_I| = |\Delta_T| + |\Delta_I| + 1.$$

Obviously, if the translations of $T$ according to the interval distribution $I$ have no common partials, then

$$|\Delta_S| = |\Delta_T| \cdot |\Delta_I|.$$

Hence, if $I$ has more than one partial, and $T$ has more than four partials, the two-level representation of the chord spectrum is more efficient than storing the whole spectrum:

$$|\Delta_I| + |\Delta_T| + 1 < |\Delta_I| \cdot |\Delta_T| = |\Delta_S|.$$

This case however corresponds to dissonant intervals.

Consonant intervals are characterized by common partials of constituent tones. Hence,

$$|\Delta_S| < |\Delta_T| \cdot |\Delta_I|.$$

Obviously, the more partials tone spectrum $T$ contains, the more non-coinciding partials are inherent in spectrum $S$. Therefore, the difference between the left-hand and right-hand parts of the above inequality is greatest when the number of partials per voice is smallest and intervals are most consonant, i.e. fifths or octaves.

It is easy to see that the spectrum of the fifth generated by a tone with seven harmonics contains 12 partials (the second harmonic of the upper tone falls on the third harmonic of the lower tone, and the fourth harmonic of the upper tone falls on the sixth harmonic of the lower tone).

The spectrum of the octave generated by a tone with seven harmonics contains 11 partials (even harmonics of the lower tone falls on the harmonics of the upper tone).

Obviously, the spectrum of a three-tone chord contains not less partials than the spectrum of its two tones. Consequently, the spectrum of a three-tone chord contains not less than 11 partials.

On the other hand, the complexity of our two-level representation of intervals chords (as a convolution of tone spectrum $T$ and interval distribution $I$) for harmonic tones with seven partials is given by the estimation

$$|\Delta_T| + |\Delta_I| + 1 = 7 + 2 + 1 = 10 < 11,$$

whence we obtain the statement of the theorem for two-tone intervals.

For three-tone chords the complexity of the two-level representation is given by the estimation

$$|\Delta_T| + |\Delta_I| + 1 = 7 + 3 + 1 = 11 \leq 11,$$

whence we obtain the statement of the theorem for three-tone chords. ■

## 3.7 Interpretation of the Results

Theorem 3 justifies the perception of a chord as a compound sound but not as a sound entirety. Assuming that the perception performs optimal data representation, we obtain its capacity to recognize physical causality in spectral data generation. This way we get a semantical knowledge concerning the chord, using some general principles of data processing only.

The role of logarithmic scale is also remarkable. Owing to the use of logarithm, linear patterns (tone spectra with multiplication of partial frequencies) become irreducible, and their superpositions become uniquely decomposable (by deconvolution of a chord spectra). Therefore, the role of logarithmic scales in perception can be explained as contributing to pattern separation.

The role of the insensitivity of the ear to the phase of signal is remarkable either. In fact, by virtue of the isomorphism between discrete spectra and polynomials (Lemma 1), we have the unique deconvolution of discrete spectra with spectral coefficients $a_n$ from a given number field, and also for integers, or Gaussian integers (complex numbers $a + bi$ with integer $a$ and $b$). Therefore, for discrete spectra with complex coefficients we obtain the unique deconvolution theorem, corresponding to the unique factorization theorem for polynomials over the field of complex numbers. Since by the fundamental theorem of algebra a polynomial over the field of complex numbers can be factored into polynomials of the first degree, this implies the deconvolution of a discrete spectrum into discrete spectra with two adjacent impulses. (From the formal point of view this resembles the decomposition of a digital filter into a superposition of second-order filters).

Being applied to our consideration, such a result would mean that a a chord spectrum $S$ can be factored into a convolution product of two-impulse spectra

$$S = \delta_p * I_1 * \ldots * I_n,$$

where each

$$I_i = a_i\delta_0 + b_i\delta_1 \quad (i = 1, \ldots, n)$$

is a two tone interval distribution with two closest impulses. In other words, a generative spectrum with two close impulses is translated by the minimal possible interval (according to the interval distribution with two close impulses), the resulting spectrum is translated again by the minimal possible interval, etc. We would obtain a hierarchy of "elementary blocks", minimal intervals over minimal intervals, which has nothing in common with human perception and physical causality in sound.

A similar situation arises for spectra with integral coefficients (including negative numbers). As seen from Example 1, the sensitivity of the ear to the inverse phase, corresponding to negative spectral coefficients, implies the decomposition of interval distribution of a major third into a convolution product of two distributions.

Thus, assuming the sensitivity of the ear to the phase of the signal (considering spectra with complex or integer coefficients), we obtain the impossibility to recognize chords corresponding to their generation.

Note that since the convolution is commutative with respect to its arguments,

$$S = I * T = T * I,$$

the interval distribution $I$ can be regarded as a generative tone and $T$ can be regarded as an interval distribution. Formally, both $I$ and $T$ can be considered as spectra of some signals.

It is known that the Fourier spectrum of a product of two signals equals to the convolution of the associated Fourier spectra, i.e.

$$F[s \cdot t] = F[s] * F[t], \tag{3.26}$$

where $s$ and $t$ are signals as functions of time, and $F[\cdot]$ is Fourier operator. For audio spectra (with logarithmic scale of frequencies), the convolution of two spectra doesn't correspond to the product of the associated signals, but we can still say that the associated signals interact somehow. Therefore, if a chord audio spectrum $S = I * T$, the interval distribution $I$ can be considered as a spectrum $I$ of a certain signal which interacts with the tone signal with spectrum $T$.

**Problem 1** *It is interesting to express analytically the interaction of the tone signal associated with spectrum $T$ and the signal associated with the interval distribution $I$, i.e. to formulate a property of audio spectra analogous to (3.26).*

Since an interval distribution $I$ can be matched to the signal $F^{-1}[I]$, the concepts of pitch and timbre are applicable to interval distributions as if they were tone spectra.

**Problem 2** *It is interesting to answer whether the type of chord (major, minor, etc.) with interval distribution $I$ corresponds to a certain "timbre" of the signal $F^{-1}[I]$ which is associated with $I$, and whether the chord root corresponds to the "pitch" of the signal $F^{-1}[I]$.*

The problem of determining the root note of a chord by its structure was posed as early as in the 18th century by J.-Ph. Rameau (1683–1764).

Finally, we would like to mention that the isomorphism between discrete spectra and polynomials over integers implies the applicability of polynomial division algorithms to the deconvolution of discrete spectra. However, the deconvolution method is efficient for justifying the principle of correlativity of perception theoretically, but may fail in practical applications to chord recognition. Indeed, in spectral approximations, partials can deviate not only from their values, corresponding to deviations of polynomial coefficients, but also from their frequencies, corresponding to changes of the degree of the associated polynomials. Since polynomial factorizations are very sensitive to the degree of polynomials, a slight spectral distortion may result in a considerable change of the spectral deconvolution. Consequently, spectral deconvolutions are unstable with respect to spectral distortions implying the instability of chord recognition with respect to small deviations of spectra. Therefore, the deconvolution method is not fitted well for chord recognition.

Another practical disadvantage of the deconvolution approach is that factoring real spectra requires much computing when the associated polynomials are of high degree. One can find the details of factoring polynomials by Kronecker's method in (Rédei 1967).

Therefore, our method of finding spectral representations by means of correlation analysis outlined in the previous chapter can be more stable and reliable in applications. On the other hand, note that by virtue of the isomorphism between spectra and polynomials, the correlation method may be used in order to factorize polynomials.

The unique factorization is valid for the polynomials in several variables (Birkhof & Mac Lane 1965, p. 76), as well as the Kronecker's method (van der Waerden 1953). Consequently our theoretical consideration is applicable to images which have two-dimensional spectra. However, by the same reasons, practical applications should be based on correlation analysis rather than on the factorization method.

## 3.8 Main Items of the Chapter

Summing up what has been said, let us formulate the main items of the chapter.

1. We have suggested a way of chord representation as generated by a tone spectral pattern translated along the $\log_2$-scaled frequency axis. It is formalized by a convolution of a tone spectral pattern and an interval distribution, corresponding to the chord structure.

2. The justification of chord recognition in its theoretical formulation is understood as the problem of unique deconvolution of the chord spectrum, corresponding to the chord spectrum generation. We have proved the indecomposability of spectra of harmonic tones and of interval distributions of two-tone intervals and major or minor triads. This corresponds to perceiving musical tones as entire sound objects and unambiguously recognizing intervals, major chords, and minor chords. Then we have established the unique deconvolution of spectra of two-tone intervals and major or minor chords with harmonic tones.

3. Note the role of logarithmic scale in our consideration. Owing to the use of logarithm, patterns with a linear structure (tone spectra with multiplication of partial frequencies) are non-linearly compressed, becoming irreducible. Therefore, the role of logarithmic scales in perception can be explained as providing the conditions for indecomposability of patterns. In particular, as follows from Lemma 3, the harmonic spectrum is irreducible, which meets the perception of a musical tone as an entire sound object rather than as a sound complex.

4. On the other hand, Theorem 3 explains the perception of chords as composed sounds. It is noteworthy that the optimality of the representation corresponds to the causality of spectral data generation. This way we obtain a semantical knowledge concerning the chord structure, using data processing only.

5. Note that if we assumed the model sensitivity to the phase of the signal, we couldn't prove the irreducibility of harmonic spectra and interval distributions of chords. On the contrary, we have shown that all spectra would become reducible to trivial two-impulse spectra. 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.