# Chapter 7

# Applications to Music Theory

## 7.1 Perception Correlativity and Music Theory

In the present chapter we discuss the correspondence between our model and music theory. This is important, because, on the one hand, music theory generalizes the experience of auditory perception acquired during centuries and, on the other hand, these generalizations are not influenced by any *a priori* assumptions which on the contrary are always inherent in scientific experiments.

Recall that basic assumptions of the given work have been formulated under the influence of the author's studies in music theory (see Section 1.4). In this chapter we compare the initial statements with the results obtained. One can expect that such a logical circle will not introduce anything principally new. However, refinement of the principle of correlativity of perception, its mathematical formulation, and computer modeling makes it possible to comment on music theory with an advanced understanding of the subject and complete initial observations with new conclusions.

The material of this chapter falls into two parts: Applications of our model to psychoacoustics (perception scales, relative and absolute hearing) and music theory in a proper sense (harmony, counterpoint, orchestration). Some earlier mentioned applications are discussed here as well.

In Section 7.2, "Logarithmic Scale in Pitch Perception", we adduce reasons in favor of logarithmic scale in pitch perception and insensitivity of the ear to the phase of the signal. It is supposed that these two properties are necessary for the decomposition of complex sounds into constituents, thus contributing to recognizing the causality in sound generation. In particular, owing to these two properties of hearing, musical tones are perceived as entire sound objects rather than as collections of sinusoidal partials, yet chords are perceived as built of complex tones.

In Section 7.3, "Definition of Musical Interval", we generalize the definition of musical interval to sounds with no pitch salience (as bell-like sounds). The interval is defined to be the distance of translation of a tone spectral pattern in the frequency domain. To measure this distance, the tones have not to be harmonic, and no coordinates of the tone location (fundamental frequencies) are needed. Thus the pitch is eliminated from the definition of interval. This meets the known fact that most people are capable to recognize intervals but cannot recognize pitch. Finally, we mention that timbral changes can influence on the perception of intervals and chords.

In Section 7.4, "Function of Relative Hearing", we explain that relative (interval) hearing, i.e. the capacity to perceive intervals juxtaposed to absolute hearing which is the capacity to recognize pitch, is the correlative perception in the frequency domain. Drawing analogy to vision, we suppose that interval hearing is the capacity to estimate the translation distance between similar sound objects without determining their precise location. This implies that interval hearing is based on the recognition of structure by identity and structure by similarity in audio scenes, contributing to separation and tracking simultaneous acoustical processes.

In Section 7.5, "Counterpoint and Orchestration", we discuss the difference between certain rules of counterpoint and orchestration. In particular, the use of parallel voice-leading in orchestration, which is prohibited in counterpoint, is justified as a mean for creating sonorous effects. Then we comment on such orchestration rules as using a common note for transmitting a melody from one instrument to another, using bright timbres for solo parts and soft voices for orchestral "pedals", precaution against large spacing, etc.

In Section 7.6, "Harmony", some properties of classical harmony are regarded as providing conditions for the recognizability of chords in our model, which is interpreted as providing conditions for adequate music perception. We explain a certain acoustical advantage of the scale of just intonation over the equal temperament. Then we comment on the prohibition against voice-crossing and voice-overlapping, prescription to fill in skips in voice-leading, function of ornamentation, role of bass in harmony, incompleteness of cluster chords, best recognizability of triads in the root position, and interaction between music hearing and music memory.

In Section 7.7, "Rhythm, Tempo, and Time", the three concepts enumerated are defined from the standpoint of the principle of correlativity of perception. We suppose that they characterize three different levels of correlative perception of time events. We also explain why minor changes of time data can considerably change the perception of rhythm and tempo. This follows from the instability of optima with respect to data changes.

Finally, in Section 7.8, "Summary of the Chapter", we recapitulate the main statements of the present discussion.

## 7.2 Logarithmic Scale in Pitch Perception

As follows from Chapter 3, the logarithmic scale in pitch perception together with the insensitivity of the ear to the phase of the signal predetermine some remarkable properties of perception.

Owing to logarithmic scaling, patterns with a linear structure, like tone spectra with harmonic ratio of partial frequencies $1 : 2 : \ldots : K$, being non-linearly compressed, become irreducible. This corresponds to the known fact that harmonic tones are perceived as entire sound objects rather than as compound ones.

The irreducibility of harmonic spectra is proved for power spectra, i.e. not for usual sound spectra with complex coefficients, but for spectra with real positive coefficients. This means that harmonic tones can be perceived as entire sound objects if only the ear is insensitive to the phase of the signal. Indeed, by virtue of Lemma 1 from Chapter 3, discrete spectra with complex coefficients are isomorphic to polynomials over complex numbers. By the fundamental theorem of algebra, such polynomials are always factored into polynomials of the first degree. This means that a spectrum with complex coefficients can be always decomposed into the convolution product of spectra with complex coefficients constituted by two close impulses each. In case of sound spectra, this property would imply total decomposability of all sounds (see Sections 3.5 and 3.7).

Therefore, the insensitivity of the ear to the phase of the signal is a useful property rather than an imperfectness of hearing. Together with the logarithmic scaling of pitch, it prevents from a "trivial" decomposition of sound which has no physical sense. These two properties of hearing are necessary for the decomposition of sound corresponding to different physical sources, thus contributing to the recognition of causality in the acoustical environment.

The indecomposability of harmonic tones is an important prerequisite for their use in music. Besides, harmonic tones are acoustically compatible with speech and with each other, resulting in consonant harmonies. Moreover, they are distinguishable in chords and in polyphony, and their timbral diversity is sufficiently large. Another important property of harmonic tones is the fact that high-level patterns of their relationships (chords and melodic lines) are quite stable with respect to noise, distortions, and voice changes. All of this make harmonic tones to be universal basic elements of musical compositions and a reliable carrier of semantical musical information.

Thus one can conclude that the logarithmic scale in pitch perception and the insensitivity of the ear to the phase of the signal not only contribute to the recognition of causality in sound, but also predetermine the use of tones with pitch salience as basic musical elements.

## 7.3 Definition of Musical Interval

Recall that an interval between two musical tones is usually defined as the difference between the tone pitches, or, which is equivalent, as the ratio of their fundamental frequencies (Gelfand 1981; Gut 1976). Similarly, chords and melodies are classified according to ratios of fundamental frequencies of their tones (Frances 1972). In other words, the known definitions of interval, chord, and melodic line are based on the idea of pitch.

Such definitions suppose that the perception of pitch precedes the perception of intervals. However, most people perceive intervals but fail in note identification, distinguish between major and minor chords not recognizing their roots, and sing songs transposing them arbitrarily instead of singing them in a fixed key.

These observations show that the mechanism of interval hearing is independent of the pitch recognition capability. Moreover, since most listeners recognize intervals but rather few recognize pitch, it is the interval hearing which is predominant in music perception. In this connection we formulate a definition of interval, chord, and melodic line not using the concept of pitch. Since chords and melodic lines are constituted by intervals, the definition of interval is principal, and definitions of chord and melodic line are its derivatives.

To illustrate our approach to defining intervals between tones with no reference to pitch, we draw analogy to the distance between visual objects. Let us show that in certain cases the distance between objects can be defined, even if it is difficult to define their precise location.

For example, consider a square on the plane shown in Fig. 7.1. The definition of its coordinates is not evident: They can be identified either with its bottom left angle, or with its center, or with some other point. In any case, the definition of coordinates of a complex object requires an additional convention. However, there is no other way of defining coordinates of a complex object other than by representing the object by a certain point.

The distance between objects is usually defined as the distance between their representative points, like mean points, centers of gravity (Fig. 7.2), or some others chosen by special rules, like in case of the Hausdorff distance between two sets. However, if the objects are *equal* and *equally oriented* in space then the distance can be defined with no reference to representative points but directly, by the magnitude of corresponding translation of the objects. This is illustrated in Fig. 7.3 where the idea of distance between two equal squares is quite evident. Since each point of the square is translated by the same distance $d$, all the points are equally representative, and there is no need to single out one of them in order to represent the square. The distance is measured between "entire" objects, but not between their representative points whose

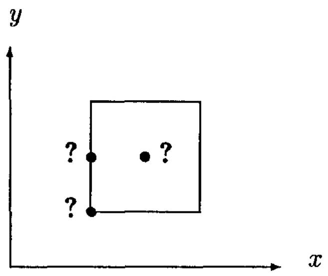


Figure 7.1: Ambiguity in defining coordinates of a complex object

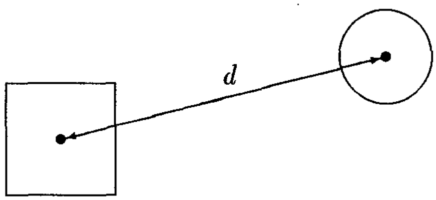


Figure 7.2: Distance between dissimilar objects

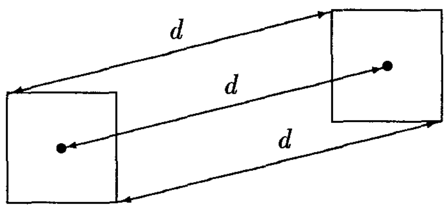


Figure 7.3: Distance between similar objects

coordinates must be determined.

Thus *if objects are equal and equally oriented in space then the distance between the objects is measured directly, without attributing the distance to some representative points of the objects.*

The need for attributing the distance to some points emerges if only the objects are dissimilar, or differently oriented in space like in Fig. 7.2. However, it should be taken into account that a reference to special points complicates the idea of distance, requiring for an additional information. According to the simplicity principle in perception, any complication should be avoided, and the estimation of distance between equal objects does provide such a possibility.

The above approach to defining the distance between equal objects with no reference to representative points can be applied to defining intervals between musical tones. Since the pitch is a kind of representative point of sound objects, the distance between similar tones can be estimated with no reference to their pitch. For example, consider two tones shown in Fig. 3.2. Since the two spectra are equal and equally oriented in "acoustical space" (due to the unidimensionality of the frequency domain), the distance between them can be measured by analogy with Fig. 7.3. For this purpose it suffices to measure the distance of translation of the spectrum, which can be done directly, independently of pitch identification. In Fig. 3.2 this translation distance is equal to five semitons, corresponding to the interval of fourth.

Therefore, if two tones have identical spectra, the distance between them can be defined as the magnitude of corresponding spectral translation. We can also extend this definition to tones which are not precisely identical but similar in spectral structure, where by similarity we understand high correlation between tone spectra. Spectra of two similar tones can be regarded as generated by the same spectral pattern which is translated along the $\log_2$-scaled frequency axis with slight distortions of its envelope (linear filtering). The translation distance is said to be the *interval* between the tones.

Although our definition of interval is restricted to *tones with similar spectral structure*, it is more general than the known one based on the distance between tone pitches. Indeed, all musical sounds (which have clear pitch salience) satisfy the condition of spectral similarity, having harmonic ratio of partial frequencies $1 : 2 : \ldots : K$. The similarity of harmonic spectra is quite evident at a logarithmic frequency axis. Even if spectral envelopes are different, the structure of partial frequencies of harmonic tones is invariant with respect to pitch translations, providing for a high correlation of the tone spectra. Therefore, two harmonic spectra can be considered as resulting from a translation of a spectrum with envelope distortions. Consequently, the definition of interval proposed is applicable to all *harmonic* tones (with a pitch salience). This means that our definition covers the case considered in the traditional definition.

Besides, our definition of interval is applicable to *inharmonic* (bell-like) sounds with no pitch salience, or even band-pass noises with similar spectra. For example, the sound of two bells has the same spectral structure, whence the translation distance between the corresponding similar spectra can be measured without pitch identification.

In a sense, a harmonic tone with unambiguous pitch is analogous to a visual object with the only evident representative point, as in case of circle with its center. An inharmonic sound with ambiguous pitch is analogous to a visual object with no evident representative point, as in case of square in Fig. 7.1. Using the pitch for the determination of intervals is analogous to measuring the distance exclusively between objects whose representative points are uniquely determined, implying that we can measure the distance between circles but not squares. This evident absurdity shows the imperfectness of traditional definition of interval.

Our definition of interval, being independent of the shape of sound objects (harmonic or inharmonic tones), is analogous to the fact that the distance in Fig. 7.3 is independent of the shape of visual objects (circles or squares). This analogy proves that our definition of interval is quite natural.

If sound spectra differ considerably, as in case of violin voice and bell sound, they cannot be regarded as generated by translations of the same spectral pattern, and our direct definition of interval is not applicable. In this case the choice of representative points is necessary in order to determine the distance between sound objects (cf. the distance between dissimilar objects illustrated in Fig. 7.2). However, since the pitch of inharmonic sounds is ambiguous, the traditional definition of interval in terms of pitch is not applicable either.

Our definition of interval implies the dependence of interval recognition on voice timbres. Imagine the following experiment: At first, an interval between two similar (inharmonic) tones is determined by a salient peak of correlation function of their spectra. Then the spectral envelopes are changed gradually, in order to amplify certain partials and suppress some others, as in experiments of Shepard (1964) and Risset (1971; 1978). Gradually transforming (filtering) the spectra in this way, one can suppress the initial peak of the correlation function and amplify another peak. This means that the initial interval becomes ambiguous. One can continue this transformation of voice spectra, making the new peak predominant, implying that the new interval becomes more salient than the former. This way timbre changes can result in interval transformations, proving the predominance of spectral cues over pitch cues in interval hearing.

By the way note that the above observation implies a possibility of continuously *modifying harmonies by timbre transformations*, similarly to pitch changes by timbre transformations in the cited experiments.

## 7.4 Function of Relative Hearing

Intervals, chords, and melodic lines are high-level patterns constituted by relationships between similar low-level patterns of tones. According to our model, revealing spectral similarity is a prerequisite for recognizing audio structure. We have shown that recognizing similarity results in a decomposition of a chord into similar components which are associated with notes. At a higher representation level, these similar spectral patterns constitute an acoustical contour recognized as a chord. In dynamics, the same perception mechanism provides joining similar sounds into acoustical trajectories recognized as melodic lines.

Therefore, relative (interval) hearing can be understood as a capacity to directly recognize the distance between tones *which are similar in spectral structure.* It is an appearance of general capability of perception to recognize structure by similarity, i.e. it is the correlative perception in audio domain.

Chords and melodic lines, as patterns of high-level, are invariant with respect to voice changes and pitch transpositions. Indeed, high-level patterns of chords and melodic lines are determined by the relationships between low-level patterns which may be not identifiable (cf. with Fig. 2.1 and Fig. 2.2 where the contour of $B$ is still recognizable even if $A$ is replaced by unknown symbol $\Pi$).

It follows that interval relationships which determine high-level patterns of chords and melodic lines are more stable with respect to different factors than low-level tone patterns characterized by pitch and timbre. Therefore, interval relationships are fitted for carring semantic musical information better than tones and pitch.

The above observation is often ignored by music theorists concerned with the standard notation and its absolute pitch. However, in the past the stuff was not provided with a pitch standard as now, and the notation has fixed relative rather than absolute pitch. Such an interval approach to notation is seen in the parts of *thorough bass, or figured bass,* where the harmony is written down by magnitudes of intervals with respect to the bass line (see Fig. 7.4).

In notation a chord can be regarded as a graphical symbol drawn by notes (cf. Fig. 2.1) which is invariant with respect to pitch transpositions. Similarly, melodic lines can be considered as trajectories invariant with respect to transpositions; thus the graphical contour of a fugue theme in different transpositions is visually recognizable in the score.

The only imperfectness of standard notation (for visually recognizing pitch transpositions) is the diatonic scale of the stuff, implying that some intervals which are equal in notation are not equal in sound. For instance, the same distance at the stuff is inherent in the following intervals: $(e_1; f_1)$ which is equal to one semitone, and $(f_1; g_1)$ which is equal to two semitones, implying that the graphical transposition of the first interval does not correspond to its transposition in sound (Fig. 7.5).

Figure 7.4: Thorough bass notation based on interval relationships (J.S.Bach's transcription of the first measures of his aria "Empfind ich Höllenangst und Pein" (Keller 1955))

Therefore, in order to make notated chords and melodic lines invariant with respect to graphical transpositions, one has to use appropriate key signatures, corresponding to the interval of transposition (Fig. 7.6).

If the staff was graduated chromatically, a transposition would mean a graphical translation of notes with respect to the stuff. A kind of such a chromatic notation is the guitar chords fingering notation where frets correspond to semitones (Fig. 7.7). In this notation pitch transpositions precisely correspond to graphical transpositions, displaying the invariable structure of a given chord type.

Finally, note that in our model the functions of relative (interval) hearing and absolute (pitch) hearing are separated. Interval hearing is necessary at the first stage of audio data processing where it contributes to the discovery of similar audio patterns and to voice separation. Absolute hearing performs pitch identification of voices at the second stage of audio data processing, after the voices have already been separated. This strong contrast between the functions of interval and pitch hearing is an exaggeration made in order to clarify our point of view.

Thus interval hearing contributes to recognizing melodic lines and chords. In a broader sense, it tracks simultaneous audio processes. Such a capacity is extremely important for orientation in the acoustical environment, and that may be the cause of dominance of interval hearing over absolute hearing developed in evolution.

Figure 7.5: Transposition in notation but not in sound

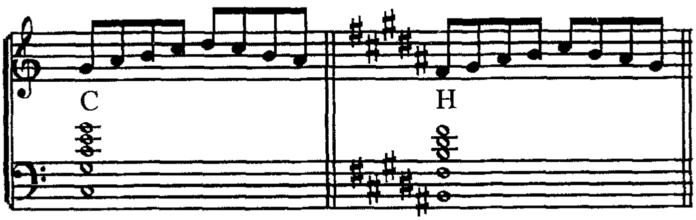


Figure 7.6: Transposed chord and melody in standard notation

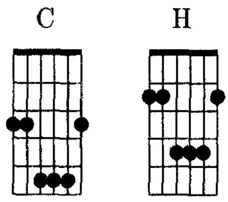


Figure 7.7: Invariance of guitar fingering with respect to chord transpositions

## 7.5 Counterpoint and Orchestration

In this section we add some remarks to the arguments from Section 1.4, concerning the prohibition against parallel primes, fifths, and octaves in the theory of counterpoint and discuss some other applications of our model to counterpoint and orchestration.

Recall that a *parallel voice-leading,* implying a parallel motion of partials of the voices, results in their fusion into a new voice with a new timbre. This effect is used in pipe organs where a single key activates several pipes tuned according to a certain chord.

The use of **timbral effect of parallel voice-leading** in orchestration is shown in Fig. 1.2. The theme played by horn is doubled by celesta at one and two octaves, and by flutes at the twelfth and the seventeenth. These five parallel parts are perceived as one voice instead of five. This way a new voice is synthesized with the first five harmonics enhanced by doubling. The idea of parallel voice-leading is emphasized by the key signatures in the score: The flute parts, corresponding to the third and fifth harmonics, are notated in G and E, respectively, while the fundamental tonality being C.

Thus the parallel voice-leading results rather in a timbral than a harmonic effect. In strict counterpoint this implies loss of a voice, reduction of harmony, and rise of a new timbre from the fusion of parallel voices. Therefore, the prohibition against parallel voices in counterpoint can be explained as preventing from breaking the homogeneity of polyphonic texture and providing the distinguishability of voices.

However, when the texture is sufficiently complex, the auditory effect may be timbral rather than harmonic, even if all the rules of voice-leading are observed. For example, in Fig. 7.8 the voices are not strictly parallel as in Fig. 1.2, but the complexity of the texture results in a fusion of the voices into a new voice with internal timbral fluctuations. If such a deviation from strictly parallel voice-leading was used in a more transparent texture, say as in Fig. 1.2, the harmonic effect would be immediately heard.

The perception of complex polyphony as a sonorousness was mentioned by Xenakis (1954; 1963) who has used this effect in order to justify stochastic composition instead of complex polyphony:

> Linear polyphony destroys itself by its very complexity; what one hears is in reality nothing but a mass of notes in various registers. The enormous complexity prevents the audience from following the intertwining of the lines and has as its macroscopic effect an irrational and fortuitous dispersion of sounds over the whole extent of the sonic spectrum. There is consequently a contradiction between the polyphonic linear system and the heard result, which is surface or mass. This contradiction inherent in polyphony will disappear

> when the independence of sounds is total. In fact, when linear combinations and their polyphonic superpositions no longer operate, what will count will be the statistical mean of isolated states and of transformations of sonic components at a given moment. The macroscopic effect can then be controlled by the mean of the movements of elements which we select. The result is the introduction of the notion of probability, which implies, in this particular case, combinatory calculus. Here, in a few words, is the possible escape from the "linear category" in musical thought.
>
> (The English translation is given by ( Xenakis 1971, p. 8)).

According to Xenakis, a complex polyphony results in the indistinguishability of voices which is perceived as sonorousness. Since the distinguishability of voices is characterized in our model by the voice separability, the model of chord recognition can be applied to **measuring the degree of sonorousness in polyphony**. A complex polyphony whose voices are not separable (not distinguishable) should be recognized as sonorousness. If parts are separable (corresponding to the distinguishability of voices) then the polyphonic origin should be recognized as predominant. Characterizing intermediate states between good separability of voices and their total fusion by the percentage of true recognition of voices, one can quantitatively estimate the ratio of polyphonic/sonorous effects.

Our model explains the predominance of dynamic perception over static perception (Bregman 1990), or the suppression of vertical (harmony) cues by linear (melodic, polyphonic) cues. For example, each harmonic vertical in Fig. 1.2, being taken separately, can be perceived as a chord. (Even partials of a tone can be distinguished in a so called analytical mode of listening.) Nevertheless, in dynamics parallel voices fuse into one with a rich timbre. In our model this follows from the fact that recognition hypotheses based on analysis of dynamic melodic intervals between a given chord and its neighbors are numerous, in contrast to the singularity of the recognition hypothesis based on analysis of static harmonic intervals. At the stage of final decision making (see Section 5.6) this implies a majority dominance of dynamic hypotheses over the only static one, meaning the dominance of linear cues over vertical ones in recognition.

In the theory of counterpoint the principle of dominance of melodic cues over harmonic cues is expressed as a **prescription to pay more attention to logical development of voices rather than to adjusting them to consonant verticals**.

Also note that the recognition of melodic intervals in melodic lines is simpler, i.e. requires less processing and memory, if all tones of a melody are played by the same voice. In this case the correlation of related spectral patterns is maximal. Conversely, if one note is played by one instrument, the following


Figure 7.8: Sonorous effect from a complex harmony in *Bolero* by M.Ravel

note by another instrument, and so on, the correlation of spectral patterns is much less, and their linking into a melody is more difficult.

Generally speaking, the continuity of timbre is very important for the correlativity of perception, since the timbral continuity makes adjacent time-cuts similar, implying their correlation and linking into acoustical processes.

In particular, this explains the known orchestration rule of **transmitting a melody from one instrument to another through a common note**: The last note played by the first instrument is prescribed to be the first note played by the second instrument. If the second instrument enters after the last note of the first instrument, the effect is recognized as starting a new phrase. This is illustrated by Webern's orchestration of *Ricercar* from J.S.Bach's *Musical Offering* (see Fig. 7.9 and 7.10). In order to provide a fine segmentation (pointillistic) effect, the instruments do not overlap. The only exception is the transmittance of melody from horn to trombone through common note $d_1$ which doesn't break the phrase.

As follows from our model, linking tones into melodic lines is more reliable if their correlation is more salient. In turn tone spectra are more correlated if there are more partials in the voice spectrum. That is an explanation of the **use of bright voices (with many partials) for solo and leading voice**. As mentioned in Section 5.5, voices with a few partials are not easily recognizable in chords. Being not structurally salient, such voices interact with each other, resulting in accidental patterns and creating a uniformly dense background. This property of voices with a few partials makes them suitable for transparent "orchestral pedals", *ripieno* parts, etc.

In this connection note that the leading voice in Fig. 1.2 and Fig. 7.8 is intentionally enriched by additional parallel parts which make the theme brighter. At the same time, the upper overtones of the theme are given to flutes with their soft timbre. This is done because these parts should blend with the main voice, not being too much salient themselves. The same reason explains why organ mixture registers are also compiled from pipes with a soft timbre.

The model explains the **prohibition against large spacing** in counterpoint, i.e. against leading voices more than an octave apart, the only exception being made for bass (Fig. 7.11). Recall that in order to avoid octave autocorrelation of voices, we restrict maximal intervals considered to 12 semitones (see Section 5.3). Under such a restriction, distant tones (spaced by intervals larger than an octave) cannot be recognized, except for the lowest tone which is always recognizable by the interval of prime (see Section 5.4).

On the other hand, a distant voice which cannot be put into correlation with other voices can be recognized only as a separate trajectory but not as a voice which interacts harmonically with other parts. Such an opposition of distant voice to others is a prerequisite for using distant voices as solo parts

Figure 7.9: Beginning of *Ricercar* from J.S.Bach's *Musical Offering*


Figure 7.10: A.Webern's orchestral arrangement of J.S.Bach's *Ricercar*

Figure 7.11: Spacing between voices more than an octave prohibited in counterpoint

but not in a homogeneous polyphony.

This is a reason for the **reservation of a free vertical space for solo** in orchestration. As follows from the above paragraph, if melodic intervals of solo part do not overlap with those of the accompaniment then the separation of solo and accompaniment becomes easier. Leading a solo part above or below other voices provides a free vertical space for solo. It is well known that soprano and bass voices are perceived better than others.

Thus some rules of orchestral arrangement can be explained from the standpoint of our model. Keeping to orchestration rules, one simplifies the perception adequate to the score. On the contrary, an infringement of these rules can result in an inadequate perception of notated parts. Certainly, one can deviate from orchestration rules intentionally in order to obtain some special audio effects.

We can conclude that the rules of counterpoint are aimed at the homogeneity of polyphonic texture, whereas the rules of orchestration are aimed at the control over certain auditory effects. The polyphonic homogeneity is only one of such effects, and that is why the rules of counterpoint are not always observed in orchestration. In a sense, orchestration rules are more general. They may be reduced to the rules of counterpoint in particular cases, but in other cases they are too restrictive.

## 7.6 Harmony

Similarly to the previous section, the model explains some rules of harmony as simplifying music perception adequate to the score, which in our model corresponds to simplifying music recognition. Since our comments are made from a rather special standpoint, they don't pretend to cover all aspects of the phenomena discussed.

| $c$ | $d$ | $e$ | $f$ | $g$ | $a$ | $h$ | $c_1$ |
|---|---|---|---|---|---|---|---|
| 1 | 9/8 | 5/4 | 4/3 | 3/2 | 5/3 | 15/8 | 2 |

Figure 7.12: Frequency ratios for the scale of just intonation in *do-major*

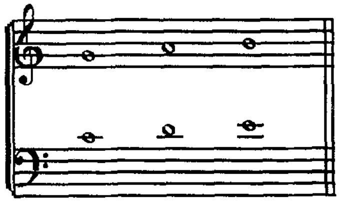

Figure 7.13: The voices parallel in notation but not in sound

The model exhibits a certain acoustical **advantage of just intonation** over the equal temperament. Recall that in the equally tempered scale all intervals equal in notation are equal in sound. It doesn't always hold in the *scale of just intonation* which is generated by octave transpositions of tones of acoustically pure tonic, dominant, and subdominant triads. The frequency ratios of diatonic degrees to the tonic in just intonation are shown in Fig. 7.12 (Rossing 1990, p.173). Note that in just intonation the fifths of tonality $C$ have the following frequency ratios:

- frequency ratio of fifth $(c; g)$ is equal to 2:3;
- frequency ratio of fifth $(d; a)$ is equal to $27 : 40 \neq 2 : 3$;
- frequency ratio of fifth $(e; h)$ is equal to 2:3.

This implies that the fifths in Fig. 7.13, looking parallel in notation, are not precisely parallel in sound. Since in our model every voice parallelism complicates their separation, the scale of just intonation, preventing from some parallelisms, provides the conditions for better voice separability. However, the difference between the two tunings mentioned, just intonation and equal temperament, is too small to be important. The perceptual effect is rather a different "color" of chords $(d; f; a)$ and $(e; g; h)$ in just intonation.

There are some other examples of correspondence between the rules of harmony and recognizability of voices in our model. Consider the **prohibition against voice crossing**, i.e. exchanging positions of two voices when a lower voice becomes the upper voice and vice versa (Fig. 7.14). If voices are not very much different, their crossing is usually perceived as one instrument playing the upper notes and another instrument playing the lower notes as if the voices were not crossed (McAdams & Bregman 1979). In our model the related recognition result is obtained if the intervals considered are restricted to small values, which requires less processing and memory (corresponding to easier perception).

The same reasons explain the **prohibition against voice overlapping.** Recall that voices are said to overlap if the lower voice moves above the former note of the upper voice, or the upper voice moves below the former note of the lower voice (Fig. 7.15).

**Filling skips** (a retrograde movement of a melody after a skip shown in Fig. 7.16) can be also justified. Indeed, filling a skip reduces the melodic interval between the chord where the skip is filled in and its second predecessor (from which the skip starts). In our model the smaller a melodic interval is, the smaller is the amount of data and processing which are required to recognize it correctly. Interpreting the simplification of recognition as the simplification of perception, we conclude that the prescription to fill skips in provides conditions for making the perception of melodies easier, i.e. making melodies more natural.

The reasons adduced here imply that it is easier to perceive a melody with small intervals (primes and seconds) than that with skips (thirds and larger intervals). In theory the corresponding voice-leading is said to be *melodic*, which emphasizes the idea of naturalness of melody, whereas a voice-leading by larger intervals is said to be *harmonic*, which indicates at some other organization of melody.

The extreme reduction of intervals between successive tones results in a **glissando** (continuous gliding from one pitch to another). As mentioned in the previous section, the principle of correlativity gives best results in recognizing continuous changes, because the correlation is greater for identical blocks of data (in the previous section we have discussed continuity in timbre; here we consider continuity in pitch). In our model a glissando is well recognizable, corresponding to easy perception. Indeed, weak sounds of Hawaiian guitar are heard distinctly even in a background of a rather loud and dense accompaniment, proving such a hypothesis.

Similarly to *glissando*, small pitch fluctuations make perceptual segregation of tones easier. Hence, we can justify the **use of ornamentation**, e.g. *vibrato*, trills, grace-notes, gliding at attacks of sounds, etc. The ornamentation, besides its coloring function, activates the perception (gives additional cues for true recognition) and attracts attention to the ornamented tones due to the recognition of their movement in a less variable background.

Such an enhancing function of ornamentation explains its wide use in harpsichord music with no other possibility to attract attention to the leading voice.

Figure 7.14: Voice crossing prohibited in counterpoint

Figure 7.15: Voice overlapping prohibited in counterpoint

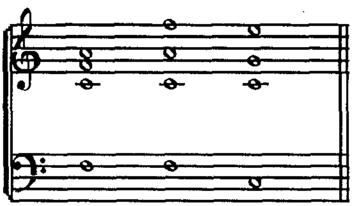

Figure 7.16: Filling skips in strict counterpoint

On the contrary, the ornamentation is not used in accompaniment which must be less salient than the leading voice. It is remarkable that ornamentation is notated by special signs, other than for melody notation. This distinction indicates at a special composition function of ornamentation, rather articulatory and intonational than structural.

The model demonstrates the **fundamental role of bass voice** in harmony. As said in Section 5.4, every tone of a chord can be recognized by its correlation with the lowest tone, because the corresponding interval is not masked by any lower parallel interval. In a sense, the bass tone determines the perception of a chord, being a reference tone for all other tones.

The model reveals the **harmonic incompleteness of cluster chords** which are constituted by equal intervals (Fig. 7.17). The lower interval, shown by the tie, is repeated several times in the chord. This means that the lower interval masks the upper parallel intervals, making them less salient. As an effect, the lower interval determines the perception of the chord, since its upper tones blend with two lower tones.

In a musical context, tones of a cluster chord can be easily separated by recognizing melodic intervals between a given chord and its neighbors. In Fig. 7.17 cluster chords are shown as suspensions with resolutions. One can see that melodic intervals between any two matched chords are not parallel, enabling tone separation by melodic intervals. Therefore, in a musical context cluster chords are quite acceptable, if only they don't form parallel sequences like in Fig. 7.18. One can see that in such a case the harmonic incompleteness of cluster chords is persistent, since all melodic intervals between successive chords are parallel.

The model explains an **easy perception of major chords in the root position** (when the root of a chord is the lowest tone). In Fig. 7.19 one can see that there is no parallel melodic interval between any two fundamental triads of major key (tonic–subdominant, subdominant–dominant, and dominant–tonic). Therefore, major triads in the root position can be correctly recognized under worst conditions (in the computer experiments described in Section 5.5 the worst conditions are the following: Accuracy of spectral representation within one semitone, harmonic voices with 16 partials, and maximal intervals restricted to 12 semitones).

The same conclusion relates to fundamental triads in natural minor, but not in harmonic minor (with major dominant). For harmonic minor, the recognition of fundamental triads is complicated by parallel melodic intervals between some of the chords shown in Fig. 7.20. This observation meets the experience of music education: The beginners write musical dictations in major key better than in minor.

We see that the rules of harmony discussed are aimed at the perception adequate to the score, which in our model corresponds to the conditions pro-

Figure 7.17: Cluster chords with resolutions

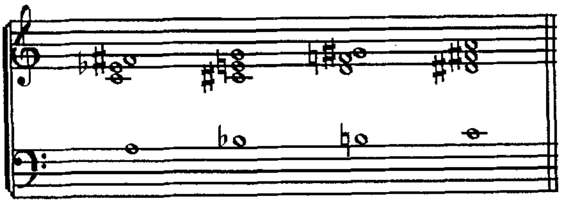

Figure 7.18: Sequence of cluster chords

Figure 7.19: Major triads in root position in harmonic and melodic junctions

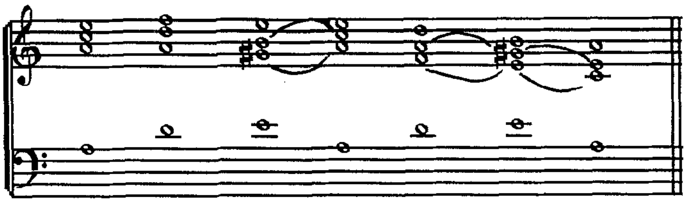

Figure 7.20: Minor triads in root position in harmonic and melodic junctions

viding true recognition. An infringement of these rules causes an inadequacy of music perception, or complicates it, corresponding to recognition mistakes or the necessity of considerable amount of processing and memory for correct recognition.

Finally, note that the **interaction between musical hearing and musical memory** can be illustrated by the model. The results of chord recognition depend greatly on the accuracy of spectral representation, on the size of maximal intervals considered, as well as on the number of chords or time-cuts confronted, i.e. on the duration of the musical excerpt held in the memory. Since a better recognition reliability requires more processing and memory, we can say that the better the memory is, the better is musical hearing.

## 7.7 Rhythm, Tempo, and Time

In Chapter 6 we have applied our model of correlative perception to rhythm recognition. Let us repeat briefly our main conclusions concerning the nature of rhythm, tempo, and time.

According to the principle of correlativity of perception, a sequence of time events is represented in terms of generative groups of events, associated with repetitious rhythmic patterns. Repeated or slightly distorted rhythmic patterns are considered as subjective time units for tempo tracking. The "trajectory" drawn by these reference units with respect to the time axis is associated with a high-level pattern of tempo curve. Among all possible representations of data, the representation with least total complexity is chosen, while the total complexity being shared between rhythmic patterns and tempo curve.

Thus rhythm and tempo are defined as *complementary attributes of optimal representation of time data.* Rhythmic patterns are defined to be generative units of such a representation and tempo is defined by time interrelationships between them. In our model tempo and rhythm cannot be considered separately, since the two representation levels are interdependent. Moreover, minimizing the total complexity of the representation, which is shared between the two levels, we adjust rhythm to tempo and tempo to rhythm.

As shown in Section 2.2 and Chapter 6, the same sequence of time events can be interpreted either in terms of rhythm, either in terms of tempo, or both. In example from Section 2.2, the sequence of durations is interpreted first in terms of rhythm (single complex rhythmic pattern under a constant tempo) and then it is interpreted in terms of tempo (simple rhythmic pattern repeated under a tempo change), depending on the context.

Since the choice of interpretation is guided by the idea of least complexity, the optimal representation can be unstable with respect to distortions of input data, which is caused by a general instability of optima. To illustrate it, imagine an infinity of representations (e.g. different representations of time

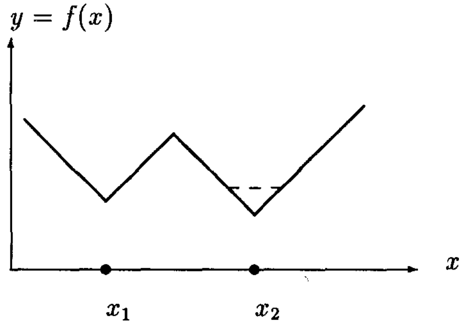


Figure 7.21: Instability of optima

events) conventionally associated with points at the axis of $x$ in Fig. 7.21. Let the complexity of representation be a function $y = f(x)$. Suppose that the graph of this complexity function looks like in Fig. 7.21. Then representation $x_2$ is least complex, being optimal. Now suppose that the initial data are changed a little, resulting in a little increase in the complexity of representations which are close to $x_2$, as shown in Fig. 7.21 by dotted line. This is sufficient to make representation $x_2$ not optimal any longer (not least complex). In Fig. 7.21 this modification results in the optimality of representation $x_1$ which is quite distant from $x_2$.

Thus a slight distortion of initial data may result in a considerable change of optimal representation, implying an alternative interpretation of time events in terms of tempo and rhythm. One can observe the instability of optimal data representation with the example illustrated by Fig. 2.3 and 2.4, where the change of optimal representation is caused not even by a change of time data, but by some external factors as placing the rhythmic progression into a melodic context.

Our approach to defining the time is also influenced by the principle of correlativity of perception. Time patterns are considered as an intermediate representation level between the level of rhythmic patterns and that of tempo curve. A time pattern is defined to be a stable pre-image with respect to elaboration of generative rhythmic patterns (cf. multilevel rhythm representation in Fig. 6.1 and with the representation obtained by the end of Section 6.8).

Summing up what has been said, we conclude that three concepts, tempo, rhythm, and time correspond to three different levels of optimal time data representation. Since we optimize the representation in whole, making it least complex, the tempo, rhythm, and time are tightly interdependent. Since they determine each other, their separate recognition is not possible in the model, and we recognize them simultaneously in their interaction.

## 7.8 Summary of the Chapter

Let us summarize the main items of this chapter.

1. The logarithmic scale in pitch perception and the insensitivity of the ear to the phase of the signal are explained as conditions which are necessary for recognizing the causality in sound generation. In particular, these two properties of hearing imply the perception of musical tones as entire sound objects rather than compound ones, as well as the perception of chords as built of tones.

2. The correlativity approach to voice separation implies a new definition of interval. The interval is defined between two tones with similar spectra as the distance of translation of the corresponding spectrum along the $\log_2$-scaled frequency axes. Therefore, we define the interval between two tones with no reference to their pitch, implying the applicability of the definition of interval to all harmonic tones and inharmonic sounds as well.

3. The nature of relative (interval) hearing is understood as the correlative perception in the audio domain. It is supposed that interval hearing is an appearance of the capacity to recognize similar sound objects and their development in time, which is necessary for voice separation and tracking simultaneous acoustical processes. Therefore, interval hearing is not a purely music perception phenomenon, but a general perception mechanism.

4. Some statements of music theory are explained as conditions which simplify the perception of polyphonic music adequate to the score. In particular, we have outlined some applications of our model to the theory of harmony, counterpoint, and orchestration. Deviations from certain rules result in special auditory effects. The consistency of statements of music theory with our model of correlative perception is interpreted as an argument in favor of its validity.

5. Three concepts related to the perception of time events, tempo, rhythm, and time, are defined from the standpoint of the principle of correlativity of perception. Rhythmic patterns are understood as reference units for tempo tracking, and the time is defined to be a stable preimage (with respect to elaboration) of generative rhythmic patterns. Tempo, rhythm, and time are defined interdependently, as certain elements of optimal representation of time data.