



UNIVERSITY OF CALIFORNIA PRESS
JOURNALS + DIGITAL PUBLISHING

Tone and Voice: A Derivation of the Rules of Voice-Leading from Perceptual Principles

Author(s): David Huron

Source: *Music Perception: An Interdisciplinary Journal*, Vol. 19, No. 1 (Fall 2001), pp. 1-64

Published by: [University of California Press](#)

Stable URL: <http://www.jstor.org/stable/10.1525/mp.2001.19.1.1>

Accessed: 16/05/2011 04:14

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=ucal>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



University of California Press is collaborating with JSTOR to digitize, preserve and extend access to *Music Perception: An Interdisciplinary Journal*.

<http://www.jstor.org>

Tone and Voice: A Derivation of the Rules of Voice-Leading from Perceptual Principles

DAVID HURON

School of Music, Ohio State University

The traditional rules of voice-leading in Western music are explicated using experimentally established perceptual principles. Six core principles are shown to account for the majority of voice-leading rules given in historical and contemporary music theory tracts. These principles are treated in a manner akin to axioms in a formal system from which the traditional rules of voice-leading are derived. Nontraditional rules arising from the derivation are shown to predict formerly unnoticed aspects of voice-leading practice. In addition to the core perceptual principles, several auxiliary principles are described. These auxiliary principles are occasionally linked to voice-leading practice and may be regarded as compositional “options” that shape the music-making in perceptually unique ways. It is suggested that these auxiliary principles distinguish different types of part writing, such as polyphony, homophony, and close harmony. A theory is proposed to account for the aesthetic origin of voice-leading practices.

Received November 20, 1996, accepted June 25, 2001

IN the training of Western musicians, the canon of harmony and voice-leading rules has been considered one of the essential foundations of the art-music craft. Of course not all composers accede to the norms of traditional harmony or voice-leading. Nor should they. Rules ought to be followed only when the composer agrees that the goal or goals embodied by the rules form worthy musical objectives, and when the rules themselves provide an effective means by which the goal(s) may be achieved. We can, therefore, ask three questions of any procedural rule: (1) What goal is served by following the rule? (2) Is the goal worthwhile? and (3) Is the rule an effective way of achieving the purported goal?

The practice of Western harmony and voice-leading has been the subject of extensive theoretical attention and elucidation (Aldwell & Schachter,

Address correspondence to David Huron, School of Music, Ohio State University, 1866 College Rd., Columbus, OH 43210.

ISSN: 0730-7829. Send requests for permission to reprint to Rights and Permissions, University of California Press, 2000 Center St., Ste. 303, Berkeley, CA 94704-1223.

1989; Berardi, 1681; Fétis, 1840; Fux, 1725; Hindemith, 1944; Horwood, 1948; Keys, 1961; Morris, 1946; Parncutt, 1989; Piston, 1978; Rameau, 1722; Riemann, 1903; Schenker, 1906; Schoenberg, 1911/1978; Stainer, 1878; and many others¹). Over the centuries, theorists have generated a wealth of lucid insights pertaining to harmony and voice-leading. Earlier theorists tended to view the voice-leading canon as a set of fixed or inviolable universals. More recently, theorists have suggested that the voice-leading canon can be usefully regarded as descriptive of a particular, historically encapsulated, musical convention. A number of twentieth-century theorists have endeavored to build harmonic and structural theories on the foundation of voice-leading.

Traditionally, the so-called “rules of harmony” are divided into two broad groups: (1) the rules of harmonic progression and (2) the rules of voice-leading. The first group of rules pertains to the choice of chords, including the overall harmonic plan of a work, the placement and formation of cadences, and the moment-to-moment succession of individual chords. The second group of rules pertains to the manner in which individual parts or voices move from tone to tone in successive sonorities. The term “voice-leading” originates from the German *Stimmführung*, and refers to the direction of movement for tones within a single part or voice. A number of theorists have suggested that the principal purpose of voice-leading is to create perceptually independent musical lines. This article develops a detailed exposition in support of this view.

In this article, the focus is exclusively on the practice of voice-leading. No attempt will be made here to account for the rules of harmonic progression. The goal is to explain voice-leading practice by using perceptual principles, predominantly principles associated with the theory of auditory stream segregation (Bregman, 1990; McAdams & Bregman, 1979; van Noorden, 1975; Wright, 1986; Wright & Bregman, 1987). As will be seen, this approach provides an especially strong account of Western voice-leading practice. Indeed, in the middle section of this article, a derivation of the traditional rules of voice-leading will be given. At the end of the article, a cognitive theory of the aesthetic origins of voice-leading will be proposed.

Understood in its broader sense, “voice-leading” also connotes the dynamic quality of tones leading somewhere. A full account of voice-leading would entail a psychological explanation of how melodic expectations and implications arise and would also explain the phenomenal experiences of tension and resolution. These issues have been addressed periodically throughout the history of music theory and remain the subject of continu-

1. No list of citations here could do justice to the volume and breadth of writings on this subject. This list is intended to be illustrative by including both major and minor writers, composers and noncomposers, didactic and descriptive authors, different periods and nationalities.

ing investigation and theorizing (e.g., Narmour, 1991, 1992). Many fundamental issues remain unresolved (see von Hippel & Huron, 2000), and so it would appear to be premature to attempt to explain the expectational aspects of tone successions. Consequently, the focus in this article will be limited to the conventional rules of voice-leading.

The purpose of this article is to address the three questions posed earlier, namely, to identify the goals of voice-leading, to show how following the traditional voice-leading rules contributes to the achievement of these goals, and to propose a cognitive explanation for why the goals might be deemed worthwhile in the first place.

Many of the points demonstrated in this work will seem trivial or intuitively obvious to experienced musicians or music theorists. It is an unfortunate consequence of the pursuit of rigor that an analysis can take an inordinate amount of time to arrive at a conclusion that is utterly expected. As in scientific and philosophical endeavors generally, attempts to explicate “common sense” will have an air of futility to those for whom the origins of intuition are not problematic. However, such detailed studies can unravel the specific origins from which intuitions arise, contribute to a more precise formulation of existing knowledge, and expose inconsistencies worthy of further study.

Plan

My presentation will proceed according to the following plan. First, the principal rules of traditional voice-leading will be reviewed. Following this review, 10 principles of auditory perception will be described and their musical implications identified. The perceptual principles described will initially be limited to a core set of 6 principles that are most pertinent to understanding voice-leading. These principles can be treated in a manner akin to axioms in a formal system from which a set of propositions can be derived.

In Part II, we will attempt to account for the traditional rules of voice-leading. In the process of the derivation, several novel rules will arise that are not normally found in theoretical writings on voice-leading. These novel rules can be regarded as theory-derived predictions. For these novel rules, we will examine compositional practice to determine whether composers typically write in a manner consistent with these additional rules. To foreshadow the results, we will see that voice-leading practices are indeed consistent with several nontraditional rules predicted by the perceptual principles.

Part III presents four additional perceptual principles that are occasionally linked to the practice of Western voice-leading. In essence, these auxil-

ary principles constitute “options” that shape the music-making in perceptually unique ways. Composers can elect to include or exclude one or more of these auxiliary principles depending on the implied perceptual goal. Once again, these principles can be treated in a manner akin to axioms in formal logic and selectively added to the core group of six central principles. Depending on the choice of auxiliary principles, we will see that alternative voice-leading systems arise. It will be argued that the choice of such auxiliary principles is one of the hallmarks of musical genres; these auxiliary principles shed light on such matters as the distinction between homophonic and polyphonic voice-leading and such unique genres as close harmony (such as the voice-leading found in barbershop quartets).

In the concluding section, a psychological account will be advanced whose goal is to explain why conventional voice-leading might be experienced by many listeners as pleasing. A number of testable predictions arise from this account. The article closes by identifying a number of unresolved issues and posing questions for further research.

Note that the purpose of this article is not somehow to “vindicate” or otherwise act as an apologist for the traditional rules of Western harmony. Nor is the intent to restrict in any way the creative enterprise of musical composition. If a composer chooses a particular goal (either implicitly or explicitly), then there are frequently natural consequences that may constrain the music making in such a way as to make the goal achievable. Musically pertinent goals might include social, political, historical, formal, perceptual, emotional, cognitive, and/or other objectives. The purpose of this article is merely to identify and clarify some of the perceptual and cognitive aspects that shape music making. In pursuing this analysis, readers should not assume that other musically pertinent goals are somehow less important or irrelevant in understanding music.

Part I: Perceptual Principles and Western Voice-Leading

RULES OF VOICE-LEADING REVIEWED

We may begin by reviewing briefly the traditional rules of voice-leading. Of course there is no such thing as “the” rules of voice-leading. Theoretical and pedagogical treatises on harmony differ in detail. Some texts identify a handful of voice-leading rules, whereas other texts define dozens of rules and their exceptions. Different pedagogical works emphasize 16th-century modal counterpoint (e.g., Gauldin, 1985; Jeppesen, 1939/1963), 18th-century Bach-style counterpoint (e.g., Benjamin, 1986; Parks, 1984; Trythall, 1993), or harmony (e.g., Aldwell & Schachter, 1989; Piston, 1978). Despite this variety, there is a core group of some dozen-odd rules that ap-

pears in nearly every pedagogical work on voice-leading. These core voice-leading rules include the following:

1. *Registral Compass Rule.* Harmony texts begin by specifying the pitch range for harmonic writing. Students should write in the region between F_2 and G_5 . Although this region corresponds to the combined pitch range for typical male and female voices, it is noteworthy that students of harmony are encouraged to write in this region, even when writing purely instrumental music.
2. *Textural Density Rule.* Harmony should be written using three or more concurrent “parts” or “voices.” The most common harmonic writing employs four voices set in overlapping pitch regions or tessituras: soprano, alto, tenor, and bass.
3. *Chord Spacing Rule.* In the spacing of chordal tones, no more than an octave should separate the soprano and alto voices. Similarly, no more than an octave should separate the alto and tenor voices. In the case of the bass and tenor voices, however, no restriction is placed on the distance separating them.
4. *Avoid Unisons Rule.* Two voices should not share the same concurrent pitch.
5. *Common Tone Rule.* Pitches common to consecutive chords should be retained in the same voice or part.
6. *Nearest Chordal Tone Rule.* If a part cannot retain the same pitch in the next sonority, the part should move to the nearest available pitch.
7. *Conjunct Motion Rule.* When a part must change pitch, the preferred pitch motion should be by diatonic step. Sometimes this rule is expressed in reverse:
8. *Avoid Leaps Rule.* Large melodic intervals should be avoided.
9. *Part-Crossing Rule.* Parts should not cross with respect to pitch.
10. *Part Overlap Rule.* No part should move to a pitch higher than the immediately preceding pitch in an ostensibly higher part. Similarly, no part should move to a pitch lower than the immediately preceding pitch in an ostensibly lower part.
11. *Parallel Unisons, Fifths, and Octaves Rule.* No two voices should move in parallel octaves, fifths, or unisons. In the case of many theorists, a more stringent version of this rule is used:
12. *Consecutive Unisons, Fifths, and Octaves Rule.* No two voices should form consecutive unisons, octaves, fifteenths (or any combination thereof)—whether or not the parts move in parallel. Also, no two voices should form consecutive fifths, twelfths, nineteenths (or any combination thereof).

13. *Exposed (or Hidden or Direct) Octaves (and Fifths) Rule.* Unisons, perfect fifths, octaves, twelfths, and so on should not be approached by similar motion between two parts unless at least one of the parts moves by diatonic step. Most theorists restrict this rule to the case of approaching an octave (fifteenth, etc.), whereas some theorists also extend this injunction to approaching perfect fifths (twelfths, etc.). Most theorists restrict this rule to voice-leading involving the bass and soprano voices; in other cases, exposed intervals between any pair of voices are forbidden.

Note that several traditional voice-leading rules are not included in the preceding summary. Most notably, the innumerable rules pertaining to chordal tone doubling (e.g., avoid doubling chromatic or leading tones) are absent. Also missing are the rules prohibiting movement by augmented intervals and the rule prohibiting false relations. No attempt will be made here to account for these latter rules (see Huron, 1993b).

Over the centuries, a number of theorists have explicitly suggested that voice-leading is the art of creating independent parts or voices. In the ensuing discussion, we will take these suggestions literally and link voice-leading practice to perceptual research concerning the formation of auditory images and independent auditory streams.

PERCEPTUAL PRINCIPLES

Having briefly reviewed the voice-leading canon, we can now identify and define the pertinent perceptual principles. In summary, the following six principles will be discussed: (1) toneness, (2) temporal continuity, (3) minimum masking, (4) tonal fusion, (5) pitch proximity, and (6) pitch co-modulation. Later, we will examine four auxiliary principles that contribute to the differentiation of musical genres: (7) onset synchrony, (8) limited density, (9) timbral differentiation, and (10) source location.

1. Toneness Principle

Sounds may be regarded as evoking perceptual “images.” An auditory image may be defined as the subjective experience of a singular sonic entity or acoustic object. Such images are often (although not always) linked to the recognition of the sound’s source or acoustic origin. Hence certain bands of frequencies undergoing a shift of phase may evoke the image of a passing automobile.

Not all sounds evoke equally clear mental images. In the first instance, aperiodic noises evoke more diffuse images than do pure tones. Even when

noise bands are widely separated and masking is absent, listeners are much less adept at identifying the number of noise bands present in a sound field than an equivalent number of pure tones—provided the tones are not harmonically related (see below).

In the second instance, certain sets of pure tones may coalesce to form a single auditory image—as in the case of the perception of a complex tone. For complex tones, the perceptual image is strongly associated with the *pitch* evoked by a set of spectral components. One of the best demonstrations of this association is to be found in phenomenon of residue pitch (Schouten, 1940). In the case of a residue pitch, a complex harmonic tone whose fundamental is absent nevertheless evokes a single notable pitch corresponding roughly to the missing fundamental.

In the third instance, tones having inharmonic partials tend to evoke more diffuse auditory images. Inharmonic partials are more likely to be resolved as independent tones, and so the spectral components are less apt to cohere and be perceived as a single sound. In a related way, inharmonic tones also tend to evoke competing pitch perceptions, such as in the case of bells. The least ambiguous pitches are evoked by complex tones whose partials most closely approximate a harmonic series.

In general, pitched sounds are easier to identify as independent sound sources, and sounds that produce the strongest and least ambiguous pitches produce the clearest auditory images. Pitch provides a convenient “hanger” on which to hang a set of spectral components and to attribute them to a single acoustic source. Although pitch is typically regarded as a subjective impression of highness or lowness, the more fundamental feature of the phenomenon of pitch is that it provides a useful internal subjective label—a perceptual handle—that represents a collection of partials likely to have been evoked by a single physical source. Inharmonic tones and noises can also evoke auditory images, but they are typically more diffuse or ambiguous.

Psychoacousticians have used the term *tonality* to refer to the clarity of pitch perceptions (ANSI, 1973); however, this choice of terms is musically unfortunate. Inspired by research on pitch strength or pitch salience carried out by Terhardt and others, Parncutt (1989) proposed the term *tonalness*; we will use the less ambiguous term *toneness* to refer to the clarity of pitch perceptions. For pure tones, toneness is known to change with frequency. Frequencies above about 5000 Hz tend to sound like indistinct “sizzles” devoid of pitch (Attneave & Olson, 1971; Ohgushi & Hato, 1989; Semal & Demany, 1990).² Similarly, very low frequencies sound like

2. More precisely, pure tones above 5 kHz appear to be devoid of pitch chroma, but continue to evoke the perception of pitch height. See Semal and Demany (1990) for further discussion.

diffuse rumblings. Frequencies in the middle range of hearing give rise to the most well-defined pitch sensations—that is, they exhibit high toneness.

The clarity of pitch perceptions has been simulated systematically in a model of pitch formulated by Terhardt, Stoll, and Seewann (1982a, 1982b). For both pure and complex tones, the model calculates a *pitch weight*, which may be regarded as an index of the pitch's clarity, and therefore, a measure of toneness. For pitches evoked by pure tones (so-called “spectral pitches”), sensitivity is most acute in the *spectral dominance region*—a broad region centered near 700 Hz. Pitches evoked by complex tones (so-called “virtual pitches”) typically show the greatest pitch weight when the evoked or actual fundamental lies in a broad region centered near 300 Hz—roughly D_4 immediately above middle C (Terhardt, Stoll, Schermbach, & Parncutt, 1986).

Figure 1 shows changes of pitch weight versus pitch for several sorts of complex tones (see also Huron & Parncutt, 1992). The pitch weights are calculated according to the Terhardt-Stoll-Seewann model. The solid curve

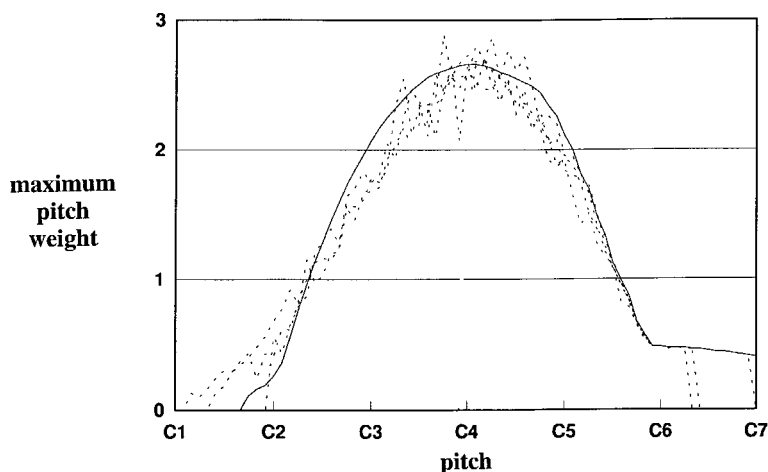


Fig. 1. Changes of maximum pitch weight versus pitch for complex tones from various natural and artificial sources; calculated according to the method described by Terhardt, Stoll, and Seewann (1982a, 1982b). Solid line: pitch weight for tones from C_1 to C_7 having a sawtooth waveform (all fundamentals at 60 dB SPL). Dotted lines: pitch weights for recorded tones spanning the entire ranges for harp, violin, flute, trumpet, violoncello, and contrabassoon. Pitch weights were calculated for each tone where the most intense partial was set at 60 dB. For tones below C_6 , virtual pitch weight is greater than spectral pitch weight; above C_6 , spectral pitch predominates—hence the abrupt change in slope. The figure shows that changes in spectral content have little effect on the region of maximum pitch weight. Note that tones having a pitch weight greater than one unit range between about E_2 and G_5 —corresponding to the pitch range spanned by the treble and bass staves. Recorded tones from Opolko and Wapnick (1989). Spectral analyses kindly provided by Gregory Sandell (Sandell, 1991a).

shows the calculated pitch weight of the most prominent pitch for sawtooth tones having a 60 dB SPL fundamental and ranging from C_1 to C_7 . The dotted curves show changes of calculated pitch weight for recorded tones spanning the entire ranges for several orchestral instruments including harp, violin, flute, trumpet, violoncello, and contrabassoon. Although spectral content influences virtual pitch weight, Figure 1 shows that the region of maximum pitch weight for complex tones remains quite stable. Notice that complex tones having a pitch weight greater than one unit on Terhardt's scale range between about E_2 and G_5 (or about 80 to 800 Hz). Note that this range coincides very well with the range spanned by the bass and treble staves in Western music.

The origin of the point of maximum virtual pitch weight remains unknown. Terhardt has suggested that pitch perception is a learned phenomenon arising primarily from exposure to harmonic complex tones produced typically by the human voice. The point of maximum virtual pitch weight is similar to the mean fundamental frequency for women and children's voices. The question of whether the ear has adapted to the voice or the voice has adapted to the ear is difficult to answer. For our purposes, we might merely note that voice and ear appear to be well co-adapted with respect to pitch.

In Huron and Parncutt (1992), we calculated the average notated pitch in a large sample of notes drawn from various musical works, including a large diverse sample of Western instrumental music, as well as non-Western works including multipart Korean and Sino-Japanese instrumental works. The average pitch in this sample was found to lie near $D\sharp_4$ —a little more than a semitone above the center of typical maxima for virtual pitch weight. This coincidence is especially evident in Figure 2, where the average notated pitch is plotted with respect to three scales: frequency, log frequency, and virtual pitch weight. The virtual pitch weight “scale” shown in Figure 2 was created by integrating the curve for the sawtooth wave plotted in Figure 1—that is, spreading the area under this curve evenly along the horizontal axis. For each of the three scales shown in Figure 2, the lowest (F_2) and highest (G_5) pitches used in typical voice-leading are also plotted. As can be seen, musical practice tends to span precisely the region of maximum virtual pitch weight. This relationship is consistent with the view that clear auditory images are sought in music making.

Once again, the causal relationship is difficult to establish. Musical practice may have adapted to human hearing, or musical practice may have contributed to the shaping of human pitch sensitivities. Whatever the causal relationship, we can note that musical practice and human hearing appear to be well co-adapted. In short, “middle C” truly is near the middle of something. Although the spectral dominance region is centered more than an octave away, middle C is very close to the center of the region of virtual

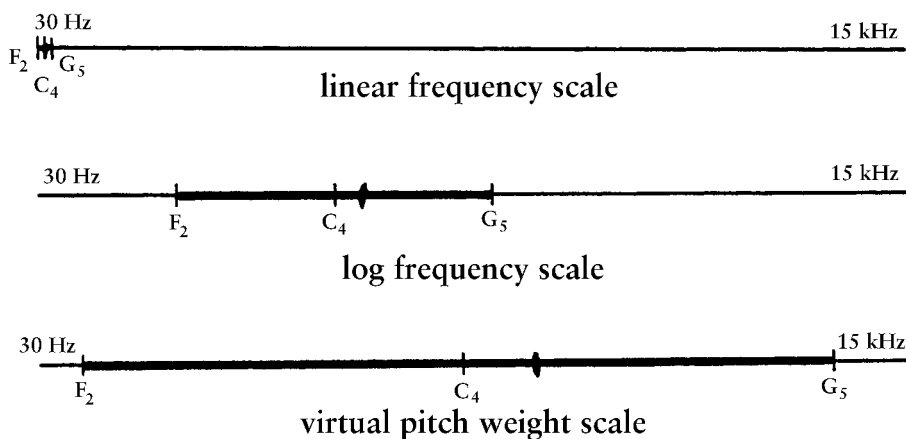


Fig. 2. Relationship of pitch to frequency, log frequency, and virtual pitch weight scales. Each line plots F_2 (bottom of the bass staff), C_4 (middle C), and G_5 (top of the treble staff) with respect to frequency, log frequency, and virtual pitch weight (VPW)-scale. Lines are drawn to scale, with the left and right ends corresponding to 30 Hz and 15,000 Hz, respectively. The solid dot indicates the mean pitch (approx. $D\#_4$) of a cross-cultural sample of notated music determined by Huron and Parncutt (1992). The VPW scale was determined by integrating the area under the curve for the sawtooth wave plotted in Figure 1. Musical practice conforms best to the VPW scale with middle C positioned near the center of the region of greatest pitch sensitivity for complex tones.

pitch sensitivity for complex tones. Moreover, the typical range for voice-leading (F_2 - G_5) spans the greater part of the range where virtual pitch weight is high. By contrast, Figure 2 shows that the pitch range for music-making constitutes only a small subset of the available linear and log frequency ranges.

Drawing on the extant research concerning pitch perception, we might formulate the following principle:

1. *Toneness Principle. Strong auditory images are evoked when tones exhibit a high degree of toneness. A useful measure of toneness is provided by virtual pitch weight. Tones having the highest virtual pitch weights are harmonic complex tones centered in the region between F_2 and G_5 . Tones having inharmonic partials produce competing virtual pitch perceptions, and so evoke more diffuse auditory images.*

2. Principle of Temporal Continuity

A second factor influencing the vividness of auditory images is their continuity. In 1971, Bregman and Campbell coined the term *auditory stream* to denote the perceptual experience of a single coherent sound activity that maintains its singleness and continuity with respect to time. A number of factors are known to influence the perception of auditory continuity. The

most obvious factor is temporal continuity, in which sound energy is maintained over a period of time.

Auditory images may be evoked by either real (sensory) or imagined (purely mental) processes. Two examples of purely mental auditory images can be found in echoic memory and auditory induction (Houtgast, 1971, 1972, 1973; Thurlow, 1957; Warren, Obusek, & Ackroff, 1972).³ In the case of auditory induction, Warren et al. generated stimuli in which intermittent faint sounds were alternated with louder sounds. The faint and loud sounds were contiguous but not overlapping. Nevertheless, the faint sounds tend to be perceived as a continuous background tone against which the loud sounds are heard to pulse. In this case, there is a mental disposition to continue a sound image even when it is physically absent.

Warren et al. were able to explain the origin of this phenomenon by showing that the frequency/intensity thresholds for auditory induction coincide closely with the thresholds for auditory masking. In other words, auditory induction mentally reinstates sounds that the listener would expect to be masked—even when the sounds are truly absent. With pure tones, robust auditory induction effects can be achieved for durations of up to 300 ms. In the case of noise bands, auditory induction may be achieved for durations of 20 s or more.

Auditory induction may be viewed as an involuntary form of auditory imagination. Fortunately, auditory induction is a subjective phenomenon that readily admits to empirical investigation and measurement. Other pertinent subjective experiences are not so easily studied. Listeners (musicians especially) are also cognizant of the existence of voluntary auditory imagination, in which the listener is able to form or sustain a purely mental image of some sound, such as the imagined sound of a tympani roll or the sound of a bubbling brook.

Although imagined sounds may be quite striking, in general, imagined sounds are significantly less vivid than actual sound stimuli. Moreover, even in the absence of sound, recently heard sounds are more vivid than less recently heard sounds. In short, auditory images have a tendency to linger beyond the physical cessation of the stimulus. In the case of real sounds, the evoked auditory images decay through the inexorable degradation of the short-term auditory store—what Neisser (1967) dubbed “echoic memory.” In general, the longer a sound stimulus is absent, the less vivid is its evoked image.

A number of experiments have attempted to estimate the duration of echoic memory. These measures range from less than 1 s (Treisman & Howarth, 1959) to less than 5 s (Glucksberg & Cowen, 1970). Typical

3. Thurlow referred to this phenomenon as the “auditory figure-ground effect;” Houtgast used the term “continuity effect.” We have elected to use the terminology proposed by Warren et al.

measures lie near 1 s in duration (Crowder, 1969; Guttman & Julesz, 1963; Rostron, 1974; Treisman, 1964; Treisman & Rostron, 1972). Kubovy and Howard (1976) have concluded that the lower bound for the half-life of echoic memory is about 1 s. Using very short tones (40-ms duration), van Noorden (1975, p. 29) found that the sense of temporal continuation of events degrades gradually as the interonset interval between tones increases beyond 800 ms.

On the basis of this research, we may conclude that vivid auditory images are evoked best by sounds that are either continuous or broken only by very brief interruptions.⁴ In short, sustained and recurring sound events are better able to maintain auditory images than are brief intermittent stimuli.

Drawing on the preceding literature, we might formulate the following principle:

2. Principle of Temporal Continuity. *In order to evoke strong auditory streams, use continuous or recurring rather than brief or intermittent sound sources. Intermittent sounds should be separated by no more than roughly 800 ms of silence in order to ensure the perception of continuity.*

Temporal Continuity and Musical Practice

The musical implications of this principle are readily apparent. In brief, the implications are evident (1) in the types of sounds commonly used in music making, (2) in the manner by which sound events succeed one another, (3) in the use of physical damping for unduly long sounds, and (4) in the differing musical treatment of sustained versus intermittent sounds.

In the first instance, the principle of temporal continuity is evident in the types of sounds commonly used in music making. Compared with most natural sound-producing objects, musical instruments are typically constructed so as to maximize the duration of the sounds produced—that is, to enhance the period of sustain. Most of the instruments of the Western orchestra, for example, are either blown or rubbed—modes of excitation that tend to evoke relatively long-lasting or continuous sounds. Even in the case of percussion instruments (such as the piano or the vibraphone), the history of the development of these instruments has shown a marked trend toward extending the resonant durations of the sounds produced. For example, the history of pianoforte construction has been marked by the continual increase in string tension, resulting in tones of longer duration. Similar historical developments can be traced for such instruments as tympani, gongs, and marimbas. In some cases, musicians have gone to extraordinary

4. The word “vivid” is meant to refer to the stability of an auditory image rather than its noticeability. Intermittent sounds are more likely to capture the listener’s attention, but they may be less apt to evoke coherent auditory streams.

lengths in order to maintain a continuous sound output. In the case of wind instruments, the disruptions necessitated by breathing have been overcome by such devices as bagpipes, mechanical blowers (as in pipe organs), and devices that respond to both “inhaling” and “exhaling” (e.g., the accordion, and, to a lesser extent, the harmonica). In addition, in some cultures, performance practices exist whose sole purpose is to maintain an uninterrupted sound. Most notable is the practice of “cyclic breathing”—as used, for example, in the Arab shawm (*zūrna*), and the north Australian didjeridoo. For stringed instruments, bowed modes of excitation have been widespread. In the extreme, continuous bowing mechanisms such as the mechanism used in the hurdy-gurdy have been devised. In the case of the guitar, solid-body construction and controlled electronic feedback have become popular methods of increasing the sustain of plucked strings.

Apart from the continuous character of most musical sounds, composers tend to assemble successions of tones in a way that suggests a persistent or ongoing existence. A notable (if seemingly trivial) fact is that the majority of notated tones in music are followed immediately by a subsequent tone. For example, 93% of all tones in vocal melodies by Stephen Foster and Franz Schubert are followed by another tone. The corresponding percentage for instrumental melodies exceeds 98%. The exceptions to this observation are themselves telling. Successions of tones in vocal music, for example, are periodically interrupted by rest periods that allow the singer(s) time to breathe. In most of the world’s music, the duration of these rest periods is just sufficient to allow enough time to inhale (i.e., about 1 or 2 s). (Longer rests typically occur only in the case of accompanied vocal music.) Moreover, when music making does not involve the lungs, pitch successions tend to have even shorter and fewer interruptions. Once again, as a general observation, we can note that in most of the world’s music making, there is a marked tendency to maintain a more or less continuous succession of acoustic events.

In tandem with efforts to maintain continuous sound outputs, musical practices also reveal efforts to terminate overlapping resonances. When musicians connect successions of pitches, problems can arise when each pitch is produced by a different vibrator. The dampers of the piano, for example, are used to truncate each tone’s normal decay. Typical piano performance finds the dampers terminating one tone at the same moment that the hammer engages the next tone. This practice is not limited to the piano, nor is it limited to Western music. Guitar players frequently use the palm of the hand to dampen one or more strings—and so enhance the sense of melodic continuation when switching from string to string. In the case of Indonesian gamelan music, proper performance practice requires the left hand to dampen the current resonator while the right hand strikes the subsequent resonator—a technique known as *tutupan*. In all of these examples,

the damping of physical vibrators is consistent with the goal of maintaining the illusion of a single continuous acoustic activity.

Of course music making also entails the use of brief sounds, such as sounds arising from various percussion instruments like the wood block. However, musicians tend to treat such brief sounds differently. Brief sounds tend not to be used to construct “lines of sound”—such as melodies—instead, these sounds are typically used intermittently. Those percussion instruments having the shortest tone durations are least apt to be used for melodic purposes. When instruments having rapid decays are used to perform melodies, they often employ tremolo or multiple repeated attacks, as in music for marimba or steel drums. When brief tones are produced by nonpercussion instruments (e.g., staccato), there is a marked tendency to increase the rate of successive tones. Contiguous staccato notes are seldom separated by more than 1 s of silence. In such cases, echoic memory serves to sustain the illusion of an uninterrupted line of sound.

As a general observation, we can note that in most of the world’s music making there is a marked preference for some sort of continuous sound activity; moreover, when instrumental sounds are brief in duration, such sounds are often assigned to nonmelodic musical tasks—even when the tones produced evoke clear pitches.

3. Minimum Masking Principle

In his Nobel-prize-winning research, Georg von Békésy showed that different frequencies produce different points of maximum displacement on the basilar membrane of the cochlea (Békésy, 1943/1949, 1960). Specifically, low frequencies cause the greatest displacement of the membrane near the apex of the cochlea, whereas high frequencies produce maximum displacements toward the oval window. Békésy, and later Skarstein (Kringelbotn, Gundersen, Krokstad, & Skarstein, 1979), carefully mapped this relationship by measuring the distance (in millimeters from the stapes) of the point of maximum displacement for a given frequency input.⁵ This correspondence between input frequency and place of maximum displacement of the basilar membrane is referred to as a tonotopic mapping or cochlear map.

Subsequent work by Harvey Fletcher linked the frequency-place coordinates of the Békésy cochlear map to experimental data from frequency discrimination and masking experiments (Fletcher, 1953, pp. 168–175).

5. Greenwood (1961a) has produced the following function to express this frequency-position relationship:

$$F = A(10^{ax} - k),$$

where F is the frequency (in hertz), x is the position of maximum displacement (in millimeters from the apex), and A , a , and k are constants: in homo sapiens $A = 165$, $a = 0.06$, and $k = 1.0$.

Fletcher showed that there is a close correspondence between distances along the basilar membrane and regions of masking. In pursuing this research, Fletcher defined a hypothetical entity dubbed the *critical band* to denote frequency-domain regions of roughly equivalent or proportional behavior (Fletcher, 1940). Subsequent research by Zwicker and others established critical bandwidths as an empirical rather than hypothetical construct. Most notably, Zwicker, Flottorp, and Stevens (1957) showed that distance along the basilar membrane accounts for changes in cumulative loudness as a function of the overall frequency spread of several tones or a band of noise (see also Scharf, 1961).

Greenwood (1961b, 1990) extended Fletcher's work by comparing psychoacoustic measures of critical bandwidth with the frequency-place coordinates of the Békésy-Skarstein cochlear map. Greenwood showed that there is a linear relationship, with one critical bandwidth being roughly equivalent to the distance of 1.0 mm on the basilar membrane.

Greenwood (1961b) went on to suggest that tonotopic effects might also account for an aspect of dissonance perception now referred to as *sensory dissonance*. Conceptually, the perception of consonance or dissonance is thought to be influenced by both cultural (learned) and sensory (innate) factors. Although little experimental research has addressed the cultural aspects of consonance or dissonance (Cazden, 1945), considerable research has established that sensory dissonance is intimately related to the cochlear map. Greenwood tested his hypothesis by comparing perceptual data collected by Mayer (1894) against the critical bandwidth/cochlear map. Mayer had collected experimental data for pure tones in which listeners were instructed to identify the smallest possible interval free of roughness or dissonance. For pure tones, this interval is not constant with respect to log frequency. Greenwood showed that there is an excellent fit between Mayer's (1894) measures and the critical bandwidth.⁶

The width of critical bands is roughly intermediate between a linear frequency scale and a logarithmic frequency scale for tones below about 400 Hz. For higher frequencies, the width becomes almost completely logarithmic. When measured in terms of hertz, critical bandwidths increase as frequency is increased; when measured in terms of semitones (log frequency), critical bandwidths decrease as frequency is increased. Figure 3 shows the approximate spacing of critical band distances using standard musical no-

6. Regrettably, the music perception community has overlooked the seminal work of Donald Greenwood and has misattributed the origin of the tonal-consonance/critical-band hypothesis to Drs. Plomp and Levelt (see Greenwood, 1961b; especially pp. 1351–1352). In addition, the auditory community in general overlooked Greenwood's accurate early characterization of the size of the critical band. After the passage of three decades, the revised ERB function (Glasberg & Moore, 1990) is virtually identical to Greenwood's 1961 equation—as acknowledged by Glasberg and Moore. For a survey of the relation of consonance and critical bandwidth to cochlear resolution, see Greenwood (1991).



Fig. 3. Approximate size of critical bands represented using musical notation. Successive notes are separated by approximately one critical bandwidth = roughly 1-mm separation along the basilar membrane. Notated pitches represent pure tones rather than complex tones. Calculated according to the revised equivalent rectangular bandwidth (ERB) (Glasberg & Moore, 1990).

tation (i.e., log frequency). Each note represents a pure tone; internote distances have been calculated according to the equivalent rectangular bandwidth rate (ERB) scale devised by Moore and Glasberg (1983; revised Glasberg & Moore, 1990).

Plomp and Levelt (1965) extended Greenwood's work linking the perception of sensory dissonance (which they dubbed "tonal consonance") to the critical band—and hence to the mechanics of the basilar membrane. Further work by Plomp and Steeneken (1968) replicated the dissonance/roughness hypothesis using more contemporary perceptual data. Plomp and Levelt estimated that pure tones produce maximum sensory dissonance when they are separated by about 25% of a critical bandwidth. However, their estimate was based on a critical bandwidth that is now considered to be excessively large, especially below about 500 Hz. Greenwood (1991) has estimated that maximum dissonance arises when pure tones are separated by about 30% to 40% of a critical bandwidth. For frequency separations greater than a critical band, no sensory dissonance arises between two pure tones. These findings were replicated by Kameoka and Kuriyagawa (1969a, 1969b) and by Iyer, Aarden, Hoglund, and Huron (1999).

In musical contexts, however, pure tones are almost never used; complex tones containing several harmonic components predominate. For two complex tones, each consisting of (say) 10 significant harmonics, the overall perceived sensory dissonance will depend on the aggregate interaction of all 20 pure-tone components. The tonotopic theory of sensory dissonance explains why the interval of a major third sounds smooth in the middle and upper registers, but sounds gruff when played in the bass region.

Musicians tend to speak of the relative consonance or dissonance of various interval sizes (such as the dissonance of a minor seventh or the consonance of a major sixth). Although the pitch distance between two complex tones affects the perceived sensory dissonance, the effect of interval size on dissonance is indirect. The spectral content of these tones, the amplitudes of the component partials, and the distribution of partials with respect to the critical bandwidth are the most formative factors determin-

ing sensory dissonance. Note that the characterizations of relative dissonance for fixed interval sizes given by musicians may also reflect cultural (learned) factors that have received little overt empirical attention (Cazden, 1945).

Plomp and Levelt hypothesized that in the writing of chords, composers would typically endeavor to maintain roughly equivalent amounts of sensory dissonance throughout the span of the chord. That is, they hypothesized that composers would typically avoid chords that produced disproportionately more dissonance in one or another region of the chord. A given sonority might be highly dissonant or consonant overall, but Plomp and Levelt supposed that the dissonance would typically be distributed homogeneously across a given sonority.

An alternative interpretation of this prediction may be offered without appealing to the concept of dissonance. Spectral components that lie within a critical band of each other cause mutual masking, which reduces the capacity of the auditory system to resolve or apprehend all of the sounds present. If composers were disposed to reduce the capacity for mutual masking, then it would be appropriate to space chordal tones in such a way that roughly equivalent amounts of spectral energy would fall in each critical band. Since critical bandwidths span more semitones in the low region than in the higher region, the notated pitches within a chord would show a distinctive distribution were this hypothesis true.

Plomp and Levelt carried out a study of chordal-tone spacing in two musical works: the third movement from J. S. Bach's Trio Sonata No. 2 for organ (*BWV 526*)⁷ and the third movement from Dvořák's String Quartet Op. 51 in E♭ major. Their analyses demonstrated an apparent consistency between the composers' arrangements of vertical sonorities and the tonotopic map—thus implying that critical bands significantly influence the vertical spacing of chordal tones.

In a replication study, Huron and Sellmer (1992) showed that Plomp and Levelt's demonstration was confounded by an unfortunate artifact and that their results could not be used to support their conclusion. However, using a more sophisticated inferential approach and a much larger musical sample, we went on to provide an alternative demonstration that confirmed Plomp and Levelt's original hypothesis. The effect of spectral spread on the spacing of chordal tones is not easily summarized; however, Figure 4 can be used to illustrate the phenomenon. Figure 4 shows the average spacing of notated (complex) tones for sonorities having various bass pitches from C₄ to C₂. For example, the first notated sonority in Figure 4 shows the *average* tenor, alto, and soprano pitches for a large sample of four-notes sonorities having C₄ as the bass pitch. (The specific sonorities notated in

7. Incorrectly cited as Trio Sonata No. 3 by Plomp and Levelt (1965).



Fig. 4. Average spacing of tones for sonorities having various bass pitches from C_4 to C_2 . Calculated from a large sample (>10,000) of four-note sonorities extracted from Haydn string quartets and Bach keyboard works (Haydn and Bach samples equally weighted). Bass pitches are fixed. For each bass pitch, the *average* tenor, alto, and soprano pitches are plotted to the nearest semitone. (Readers should not be distracted by the specific sonorities notated; only the approximate spacing of voices is of interest.) Note the wider spacing between the lower voices for chords having a low mean tessitura. Notated pitches represent complex tones rather than pure tones.

Figure 4 should not be interpreted literally; only the approximate spacing of voices is of interest.) As the tessitura of the chord descends, the pitch separation between the lower notes in the sonority tends to become larger. This is consistent with efforts to distribute spectral components in a roughly uniform manner across the basilar membrane. Such a pattern is consistent with both the goals of minimizing sensory dissonance and minimizing auditory masking.

As pointed out by theorist Walter Piston (1978), the spacing between chordal tones is not related to the spacing of partials in the harmonic series (as formerly supposed). The harmonic series exhibits the same interval spacing whatever the register of a chord. Yet composed chords show systematic interval changes with respect to register.

While acknowledging that the extant research is consistent with both the goals of minimizing sensory dissonance and minimizing auditory masking, for the purposes of voice-leading, we will formulate the relevant musical principle in terms of masking:

3. Minimum Masking Principle. *In order to minimize auditory masking within some vertical sonority, approximately equivalent amounts of spectral energy should fall in each critical band. For typical complex harmonic tones, this generally means that simultaneously sounding notes should be more widely spaced as the register descends.*

4. Tonal Fusion Principle

Tonal fusion is the tendency for some concurrent sound combinations to cohere into a single sound image. Tonal fusion arises most commonly when the auditory system interprets certain frequency combinations as comprising partials of a single complex tone (DeWitt & Crowder, 1987). Two factors are known to affect tonal fusion: (1) the frequency ratio of the component tones and (2) their spectral content. Tonal fusion is most probable

when the combined spectral content conforms to a single hypothetical harmonic series. This occurs most commonly when the frequencies of the component tones are related by simple integer ratios.

The pitch interval that most encourages tonal fusion is the aptly named *unison*. The second most fused interval is the octave, whereas the third most fused interval is the perfect fifth (DeWitt & Crowder, 1987; Stumpf, 1890). Following Stumpf, many music researchers have assumed that tonal fusion and tonal consonance are the same phenomenon and that both arise from simple integer frequency ratios. However, the extant psychoacoustic research does not support Stumpf's view. Bregman (1990) has noted that the confusion arises from conflating "smooth sounding" with "sounding as one." As we have seen, work by Greenwood (1961a, 1961b, 1990, 1991), Plomp and Levelt (1965), Kameoka and Kuriyagawa (1969a, 1969b), and Iyer et al. (1999) implicates critical band distances in the perception of tonal consonance or sensory dissonance. This work shows that sensory dissonance is only indirectly related to harmonicity or tonal fusion.

Whether or not tonal fusion is a musically desirable phenomenon depends on the music-perceptual goal. In Huron (1991b), it was shown that in the polyphonic writing of J. S. Bach, tonally fused harmonic intervals are avoided in proportion to the strength with which each interval promotes tonal fusion. That is, unisons occur less frequently than octaves, which occur less frequently than perfect fifths, which occur less frequently than other intervals. Of course concurrent octaves and concurrent fifths occur regularly in music, but (remarkably) they occur less frequently in polyphonic music than they would in a purely random juxtaposition of voices.

Note that this observation is independent of the avoidance of *parallel* unisons, fifths, or octaves. As simple static harmonic intervals, these intervals are actively avoided in Bach's polyphonic works. Considering the importance of octaves and fifths in the formation of common chords, their active avoidance is a remarkable feat (see Huron, 1991b).

In light of the research on tonal fusion, we may formulate the following principle:

4. Tonal Fusion Principle. *The perceptual independence of concurrent tones is weakened when their pitch relations promote tonal fusion. Intervals that promote tonal fusion include (in decreasing order): unisons, octaves, perfect fifths, ... Where the goal is the perceptual independence of concurrent sounds, intervals ought to be shunned in direct proportion to the degree to which they promote tonal fusion.*

Sensory Dissonance, Tonal Fusion, and Musical Practice

Before continuing with the fifth perceptual principle, we might briefly consider the relationship between sensory dissonance and tonal fusion—in

light of musical practice. As we have seen, sensory dissonance depends on more than just the pitch interval separating two tones. It also depends on the spectral content of the tones, as well as their tessitura. Nevertheless, pitch distance retains an indirect influence on sensory dissonance, and it is therefore possible to calculate the *average* dissonance for intervals of various sizes—such as the average dissonance of “a major sixth” (see Huron, 1994).

The solid line in Figure 5 shows the aggregate consonance values for musical intervals up to the size of an octave from perceptual experiments by Kaestner (1909). The bars show the interval prevalence in the upper two voices of J. S. Bach’s three-part Sinfonias.⁸ There is a fairly close fit between these two sets of data, suggesting that Bach tends to use various intervals in inverse proportion to their degree of sensory dissonance. There are a few notable discrepancies, however. The unison and octave intervals occur relatively infrequently in the Bach sample, although the sensory dissonance for these intervals is very low. In addition, the minor and major thirds appear to be “switched:” the consonance data would predict that major thirds would be more prevalent than minor thirds.

The origin of the discrepancies for the major and minor thirds is not known. It is possible that the relative prevalence of minor thirds may be an artifact of the interval content of common chord types. Major and minor

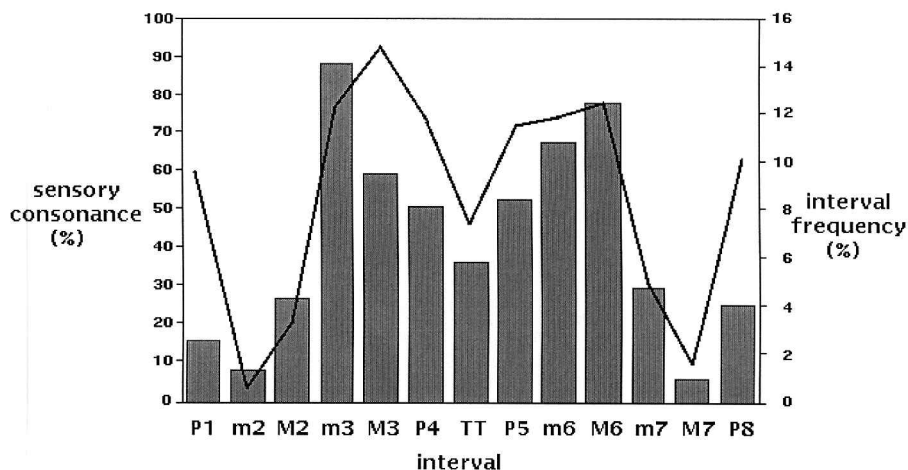


Fig. 5. Comparison of sensory consonance for complex tones (line) from Kaestner (1909) with interval prevalence (bars) in the upper two voices of J.S. Bach’s three-part Sinfonias (BWVs 787–801). Notice especially the discrepancies for P1 and P8. Reproduced from Huron (1991b).

8. Technical justifications for sampling the upper two voices of Bach’s three-part Sinfonias are detailed by Huron (1994).

triads both contain one major third and one minor third each. However, both the dominant seventh chord and the minor-minor-seventh chord contain one major third and two minor thirds. In addition, the diminished triad (two minor thirds) is more commonly used than the augmented triad (two major thirds). Hence, the increased prevalence of minor thirds compared with major thirds may simply reflect the interval content of common chords. The relative prevalences of the major and minor sixth intervals (inversions of thirds) are also consistent with this suggestion.

The discrepancies for P1 and P8 in Figure 5 are highly suggestive in light of the tonal fusion principle. We might suppose that the reason Bach avoided unisons and octaves is in order to prevent inadvertent tonal fusion of the concurrent parts. In Huron (1991b), this hypothesis was tested by calculating a series of correlations. When perfect intervals are excluded from consideration, the correlation between Bach's interval preference and the sensory dissonance Z scores is -0.85 .⁹ Conversely, if we exclude all intervals apart from the perfect intervals, the correlation between Bach's interval preference and the tonal fusion data is -0.82 . Calculating the multiple regression for both factors, Huron (1991b) found an R^2 of 0.88—indicating that nearly 90% of the variance in Bach's interval preference can be attributed to the twin compositional goals of the pursuit of tonal consonance and the avoidance of tonal fusion. The multiple regression analysis also suggested that Bach pursues both of these goals with approximately equal resolve. Bach preferred intervals in inverse proportion to the degree to which they promote sensory dissonance and in inverse proportion to the degree to which they promote tonal fusion. It would appear that Bach was eager to produce a sound that is “smooth” without the danger of “sounding as one.” In Part III, we will consider possible aesthetic motivations for this practice.

Musical Terminology: Types of Harmonic Intervals

The experimental results pertaining to sensory dissonance and tonal fusion may be used to illuminate traditional musical terminology. Music theorists traditionally distinguish three classes of harmonic intervals: *perfect consonances* (such as perfect unisons, octaves, fourths, and fifths), *imperfect consonances* (such as major and minor thirds and sixths), and *dissonances* (such as major and minor seconds and sevenths, and tritones). These interval types can be classified according to the criteria of sensory dissonance and tonal fusion. Perfect consonances typically exhibit low sensory

9. The word “preference” is not used casually here. Huron (1991b) developed an autophase procedure that permits the contrasting of distributions in order to determine what intervals are being actively sought out or actively avoided by the composer. The ensuing reported correlation values are all statistically significant.

dissonance and high tonal fusion. Imperfect consonances have low sensory dissonance and comparatively low tonal fusion. Dissonances exhibit high sensory dissonance and low tonal fusion. (There are no equally tempered intervals that exhibit high sensory dissonance and high tonal fusion, although the effect can be generated using grossly mistuned unisons, octaves, or fifths.) It would appear that the twin phenomena of sensory dissonance and tonal fusion provide a plausible account for both the traditional theoretical distinctions, as well as Bach's compositional practice.

5. Pitch Proximity Principle

In 1950, Miller and Heise observed that alternating pitches (such as trills) produce two different perceptual effects depending on the pitch distance separating the tones (Miller & Heise, 1950; see also Heise & Miller, 1951). When the tones are close with respect to pitch, quick alternations evoke a sort of "undulating effect"—like a single wavering line. However, when the pitch separation is larger, the perceptual effect becomes one of two "beeping" tones of static pitch. Musicians recognize this phenomenon as that of *pseudo-polyphony* or *compound melodic line*, in which a single sequence of pitches nevertheless evokes a sort of "yodelling" effect.

Miller and Heise's observations were replicated and extended by a number of researchers including Bozzi and Vicario (1960) and Vicario (1960), Schouten (1962), Norman (1967), Dowling (1967), van Noorden (1971a, 1971b), and Bregman and Campbell (1971). (Several of these researchers worked independently, without knowledge of previously existing work.) Of these pioneering efforts, the most significant works are those of Dowling (1967) and van Noorden (1975). Over the past three decades, however, the most sustained and significant research effort has been that of Albert Bregman (1990).

In 1975 van Noorden mapped the relationship between tempo and pitch separation on stream integration and segregation. Figure 6 summarizes van Noorden's experimental results. When the tempo is slow and/or the pitches have close proximity, the resulting sequence is always perceived as a single stream. This area is indicated in Figure 6 as region 1—below the *fission boundary* (lower line). Conversely, when the pitch distances are large and/or the tempo is fast, two streams are always perceived. This condition is indicated in Figure 6 as region 2—to the left of the *temporal coherence boundary*. Van Noorden also identified an intervening gray region, where listeners may hear either one or two streams depending on the context and the listener's disposition. Notice that the slope of the fission boundary is much shallower than the slope of the temporal coherence boundary. We will return to consider these different slopes later.

The importance of pitch proximity in stream organization is supported by a wealth of further experimental evidence. Schouten (1962) observed

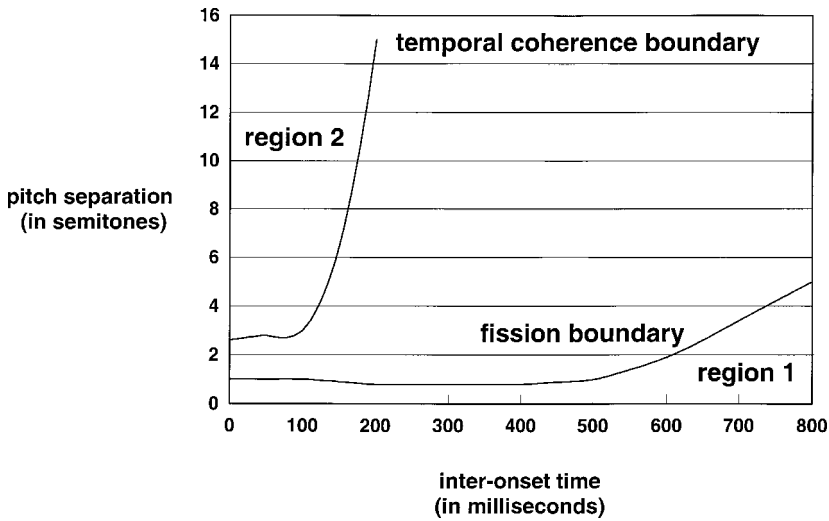


Fig. 6. Influence of interval size and tempo on stream fusion and segregation (van Noorden, 1975, p. 15). Upper curve: temporal coherence boundary. Lower curve: fission boundary. In Region 1, the listener necessarily hears one stream (small interval sizes and slow tempos). In Region 2, the listener necessarily hears two streams (large interval sizes and fast tempos).

that temporal relationships are more accurately perceived within streams than between streams — when streams are distinguished by pitch proximity alone. This work was replicated and extended by Norman (1967), Bregman and Campbell (1971), and Fitzgibbons, Pollatsek, and Thomas (1974).

In addition, Bregman and his colleagues have assembled strong evidence showing the preeminence of pitch proximity over pitch trajectory in the continuation of auditory streams. In sequences of pitches, listeners tend not to “extrapolate” future pitches according to contour trajectories. Rather they “interpolate” pitches between pitches deemed to be in the same stream (Ciocca & Bregman, 1987; Steiger & Bregman, 1981; Tougas & Bregman, 1985; summarized by Bregman, 1990, pp. 417–442). Similar effects have been observed in speech perception (Darwin & Gardner, 1986; Pattison, Gardner, & Darwin, 1986).

Further experimental work has demonstrated the perceptual difficulty of tracking auditory streams that cross with respect to pitch. Using pairs of well-known melodies, Dowling (1973) carried out experiments in which successive notes of the two melodies were interleaved. The first note of melody “A” was followed by the first note of melody “B,” followed by the second note of melody “A” followed by the second note of melody “B” etc. Dowling found that the ability of listeners to identify the melodies was highly sensitive to the pitch overlap of the two melodies. Dowling found that as the melodies were transposed so that their mean pitches diverged, recognition scores increased. The greatest increase in recognition scores

occurred when the transpositions removed all pitch overlap between the concurrent melodies.

Deutsch (1975) and van Noorden (1975) found that, for tones having identical timbres, concurrent ascending and descending tone sequences are perceived to switch direction at the point where their trajectories cross. That is, listeners are disposed to hear a “bounced” percept in preference to the crossing of auditory streams. Figure 7 illustrates two possible perceptions of intersecting pitch trajectories. Although the crossed trajectories represents a simpler Gestalt figure, the bounced perception is much more common—at least when the trajectories are constructed using discrete pitch categories (as in musical scales).

In summary, at least four empirical phenomena point to the importance of pitch proximity in helping to determine the perceptual segregation of auditory streams: (1) the fission of monophonic pitch sequences into pseudo-polyphonic percepts described by Miller and Heise and others, (2) the discovery of information-processing degradations in cross-stream temporal tasks, as found by Schouten, Norman, Bregman and Campbell, and Fitzgibbons, Pollatsek, and Thomas, (3) the perceptual difficulty of tracking auditory streams that cross with respect to pitch described by Dowling, Deutsch, and van Noorden, and (4) the preeminence of pitch proximity over pitch trajectory in the continuation of auditory streams demonstrated by Bregman et al. Stream segregation is thus strongly dependent upon the proximity of successive pitches.

On the basis of extensive empirical evidence, we can note the following principle:

5. Pitch Proximity Principle. The coherence of an auditory stream is maintained by close pitch proximity in successive tones within the stream. Pitch-based streaming is assured when pitch movement is within van Noorden’s “fission boundary” (normally 2 semitones or less for tones less than 700 ms in duration). When pitch distances are large, it may be possible to maintain the perception of a single stream by reducing the tempo.



Fig. 7. Schematic illustration of two possible perceptions of intersecting pitch trajectories. “Bounced” perceptions (right) are more common for stimuli consisting of discrete pitch sequences, when the timbres are identical.

Pitch Proximity and Music

If a musical melody is a kind of auditory stream (Dowling, 1967), then we might expect sequences of pitches to conform to the pitch proximity principle. Specifically, we would expect the majority of pitch sequences to fall below van Noorden's fission boundary and certainly to fall below the temporal coherence boundary.

The evidence concerning the prevalence of pitch proximity in music is extensive. Several authors have observed the pervasive use of small intervals in the construction of melodies, including Ortmann (1926), Merriam, Whinery and Fred (1956), and Dowling (1967). Figure 8 plots further data showing the distribution of interval sizes using samples of music from a number of cultures: American, Chinese, English, German, Hasidic, Japanese, and sub-Saharan African (Pondo, Venda, Xhosa, and Zulu). In general, the results affirm the preponderance of small intervals.

Apart from melodies, musicians know that the pitch interval between successive notes is a crucial determinant in the construction of pseudo-polyphonic passages. Dowling (1967) carried out a study of a number of Baroque solo works in order to determine the degree to which pseudo-polyphony is correlated with the use of large intervals. In measurements of interval sizes in passages deemed pseudo-polyphonic by two independent auditors, Dowling found the passages to contain intervals markedly larger than the "trill threshold" measured by Miller and Heise—a threshold similar to van Noorden's fission boundary. In a sample of pseudo-polyphonic

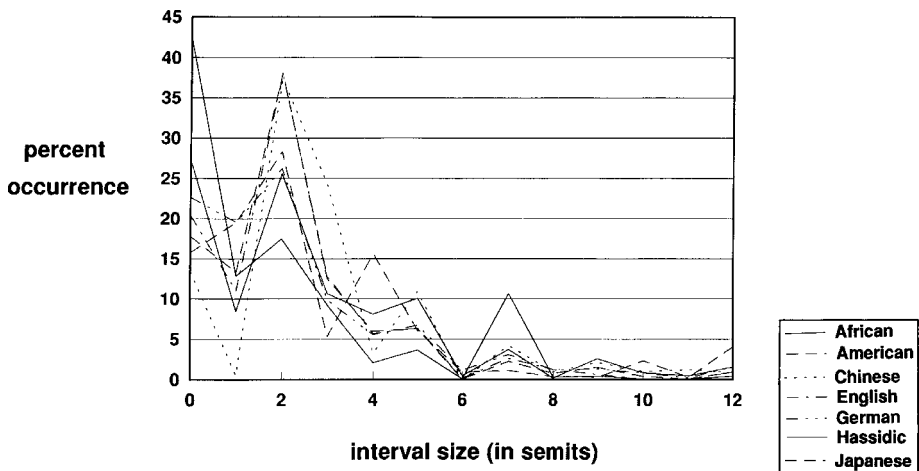


Fig. 8. Frequency of occurrence of melodic intervals in notated sources for folk and popular melodies from 10 cultures ($n = 181$). African sample includes Pondo, Venda, Xhosa, and Zulu works. Note that interval sizes correspond only roughly to equally tempered semitones.

passages by Telemann, Dowling noted that Telemann never uses intervals less than the Miller and Heise trill threshold.

As further evidence of the primacy of small intervals in non pseudo-polyphonic lines, Dowling examined the interval *preferences* of listeners. Through a series of stochastically generated stimuli, Dowling found that listeners prefer melodies that use the smaller interval sizes. Carlsen (1981), moreover, demonstrated that musicians show a marked perceptual *expectancy* for proximate pitch contours. Studying musicians from three different countries (Hungary, Germany, and the United States), Carlsen discovered that although there are significant differences between musicians in their melodic expectancies, in all cases neighboring pitch continuations of melodic stimuli are strongly favored.

Further evidence showing the consistency between musical practice and empirical research regarding pitch proximity is evident in the case of part-crossing. Notice that, except in the case of unisons, the crossing of parts with respect to pitch always violates the pitch proximity principle. No matter how the pitches are arranged, the aggregate pitch distance is always lowest when the upper voice consists of the upper pitches and the lower voice consists of the lower pitches. In a study of part-crossing in polyphonic music, Huron (1991a) showed that J. S. Bach avoids part-crossing and that he becomes most vigilant to avoid part-crossing when the number of concurrent parts is three or more.

In summary, at least five empirical observations point to the importance of pitch proximity in musical organization: (1) the preponderance of small pitch intervals in non-pseudo-polyphonic melodies observed by Ortmann, Merriam et al., and Dowling, (2) the reciprocal prevalence of large pitch intervals in pseudo-polyphonic passages found by Dowling, (3) the auditory preference for small intervals in melodies found by Dowling, (4) the perceptual expectation for small intervals in continuations of melodic contours found by Carlsen, and (5) the avoidance of part-crossing in polyphonic music measured by Huron. Musical practice thus exhibits a notable consistency with the pitch proximity principle.

Types of Melodic Intervals—The Fission Boundary

Once again, we might pause and consider the relationship between the perceptual evidence concerning pitch proximity and traditional musical terminology. Recall van Noorden's fission and temporal coherence boundaries shown in Figure 6. Not all regions in Figure 6 are equally pertinent to musical practice. Figure 9 shows distributions for note durations for samples of instrumental and vocal lines. Only durations for contiguous notated pitches have been included in this sample of 13,178 notes. Notes immediately prior to rests and at the ends of works have been eliminated from the

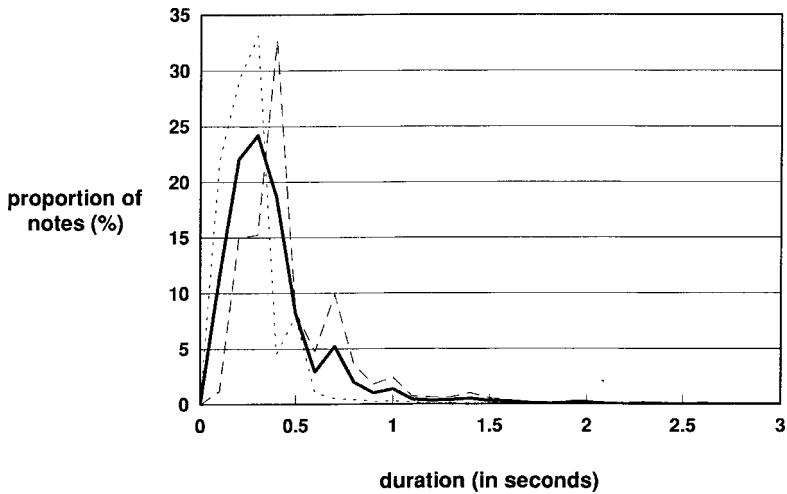


Fig. 9. Distribution of note durations in 52 instrumental and vocal works. Dotted line: note durations for the combined upper and lower voices from J. S. Bach's two-part Inventions (*BWVs* 772–786). Dashed line: note durations in 38 songs (vocal lines only) by Stephen Foster. Solid line: mean distribution for both samples (equally weighted). Note durations were determined from notated scores using tempi measured from commercially available sound recordings. Graphs are plotted using bin sizes of 100 ms, centered at the positions plotted.

sample in order to avoid the effect of final lengthening and to make the data consistent with the pitch alternation stimuli used to generate Figure 6.

A mean distribution is indicated by the solid line. In general, Figure 9 shows that the majority of musical tones are relatively short; only 3% of tones are longer than 1 s in duration. Eighty-five percent of tones are shorter than 500 ms; however, it is rare for tones to be less than 150 ms in duration. The median note durations for the samples plotted in Figure 9 are 0.22 s for instrumental and 0.38 s for vocal works.

This distribution suggests that the majority of tone durations in music are too long to be in danger of crossing the temporal coherence boundary—that is, where two streams are necessarily perceived. Notice that compared with the temporal coherence boundary, the fission boundary is more nearly horizontal (refer to Fig. 6). The boundary is almost flat for tone durations up to about 400 or 500 ms after which the boundary shows a positive slope. Figure 9 tells us that the majority of musical tones have durations commensurate with the flat portion of the fission boundary. This flat boundary indicates that intervals smaller than a certain size guarantee the perception of a single line of sound. In the region of greatest musical activity, the fission boundary approximates a constant pitch distance of about 1 semitone.

As in the case of harmonic intervals, music theorists traditionally distinguish two main classes of melodic intervals: *conjunct* or step motions, and *disjunct* or leap motions. In Western music, the dividing line between step and leap motion is traditionally placed between a major second and a minor third; that is, a major second (2 semitones) is considered a conjunct or step motion, whereas a minor third (3 semitones) is considered the smallest disjunct or leap motion. In other cultures, such as those that employ the common pentatonic scale, the maximum “step” size is roughly 3 semitones. It is plausible that the fission boundary in effect identifies a psychoacoustic basis for the distinction between conjunct and disjunct melodic motions. What theorists call conjunct intervals are virtually guaranteed to evoke the perception of stream continuation.

Melodic Motion—The Temporal Coherence Boundary

Unlike the fission boundary, the temporal coherence boundary plotted in Figure 6 shows a marked positive slope. This means that, in order for a sequence of pitches to break apart into two streams, tempo is the predominant factor—although pitch interval continues to play a significant role. If the temporal coherence boundary influences musical organization, we might expect to see tradeoffs between pitch interval and instantaneous tempo. For example, we might predict that large pitch leaps would be associated with tones of long duration.

Van Noorden (1975, p. 48) and Shepard (1981, p. 319) independently drew attention to the similarity between Miller and Heise’s trill results and K rte’s third law of apparent motion in vision (K rte, 1915). K rte carried out a number of experiments using two lamps that could be alternately switched on and off. K rte found that the sense of apparent motion depends on the distance separating the two lamps and their speed of switching. If the lamps are placed farther apart, then the rate of switching must be reduced in order to maintain a sense of apparent motion between the lamps. If the switching rate is too fast, or the lamps are placed especially far apart, then the viewer sees two independent flickering lights with no sense of intervening motion. A possible explanation for this loss of apparent motion is that it is implausible for a single real-world object to move in accordance with the presumed trajectory.

The parallel between K rte’s results and Miller and Heise’s trills is obvious and direct. It would seem that the sense of continuation between two tones is an auditory analog to apparent motion in vision. Note that both K rte’s results and Miller and Heise’s results pertain to *perception*. Research by Huron and Mondor (1994) suggests that the cognitive basis for this parallel can be attributed to the real-world generation or production of motion.

In Huron and Mondor (1994), it was argued that both K rte’s third law of apparent motion and Miller and Heise’s results arise from the kinematic principle known as Fitts’ law (Fitts, 1954). Fitts’ law applies to all muscle-powered or autonomous movement. The law is best illustrated by an example (refer to Figure 10). Imagine that you are asked to alternate the point of a stylus as rapidly as possible back and forth between two circular targets. Fitts’ law states that the speed with which you are able to accomplish this task is proportional to the size of the targets and inversely proportional to the distance separating them. The faster speeds are achieved when the targets are large and close together—as in the lower pair of circles in Figure 10.

Because Fitts’ law applies to all muscular motion, and because vocal production and instrumental performance involve the use of muscles, Fitts’ law also constrains the generation or production of sound. Imagine, for example, that the circular targets in Figure 10 are arranged vertically rather than horizontally. From the point of view of vocal production, the distance separating the targets may be regarded as the pitch distance between two tones. The size of the targets represents the pitch accuracy or intonation. Fitts’ law tells us that, if the intonation remains fixed, then vocalists will be unable to execute wide intervals as rapidly as for small intervals.

In Huron and Mondor (1994), both auditory streaming and melodic practice were examined in light of Fitts’ law. In the first instance, two perceptual experiments showed that increasing the variability of pitches (the auditory equivalent of increased target size) caused a reduction in the fission of auditory streams for rapid pitch alternations—as predicted by Fitts’

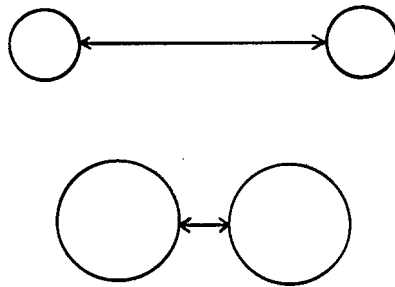


Fig. 10. Two pairs of circular targets illustrating Fitts’ law. The subject is asked to alternate the point of a stylus back-and-forth as rapidly as possible between two targets. The minimum duration of movement between targets depends on the distance separating the targets as well as target size. Hence, it is possible to move more rapidly between the lower pair of targets. Fitts’ law applies to all muscle motions, including the motions of the vocal muscles. Musically, the distance separating the targets can be regarded as the pitch distance between two tones, whereas the size of the targets represents pitch accuracy or intonation. Fitts’ law predicts that if the intonation remains fixed, then vocalists will be unable to execute wide intervals as rapidly as for small intervals.

law. With regard to the perception of melody, Huron and Mondor noted that empirical observations of preferred performance practice (Sundberg, Askenfelt & Frydén, 1983) are consistent with Fitts' law. When executing wide pitch leaps, listeners prefer the antecedent tone of the leap to be extended in duration. In addition, a study of the melodic contours in a cross-cultural sample of several thousand melodies showed them to be consistent with Fitts' law. Specifically, as the interval size increases, there is a marked tendency for the antecedent and consequent pitches to be longer in notated or performed duration. This phenomenon accords with van Noorden's temporal coherence boundary. In short, when large pitch leaps are in danger of evoking stream segregation, the instantaneous tempo for the leap is reduced. This relationship between pitch interval and interval duration is readily apparent in common melodies such as "My Bonnie Lies Over the Ocean" or "Somewhere Over the Rainbow." Large pitch intervals tend to be formed using tones of longer duration—a phenomenon we might dub "leap lengthening."

It would appear that both Korte's third law of apparent motion and the temporal coherence boundary in auditory streaming can be attributed to the same cognitive heuristic: Fitts' law. The brain is able to perceive apparent motion, only if the visual evidence is consistent with how motion occurs in the real world. Similarly, an auditory stream is most apt to be perceived when the pitch-time trajectory conforms to how real (mechanical or physiological) sound sources behave. In short, it may be that Fitts' law provides the origin for the common musical metaphor of "melodic motion" (see also Gjerdingen, 1994).

Further evidence in support of a mechanical or physiological origin for pitch proximity has been assembled by von Hippel (2000). An analysis of melodies from cultures spanning four continents shows that melodic contours can be modeled very well as arising from two constraints—dubbed *ranginess* and *mobility*. Such a model of melodic contour is able to account for a number of melodic phenomena, notably the observation that large leaps tend to be followed by a reversal in melodic direction.

6. Pitch Co-modulation Principle

As early as 1863, Helmholtz suggested that similar pitch motion contributes to the perceptual fusion of concurrently sounded tones. In recent decades, Helmholtz's suggestion has received considerable empirical confirmation and extension. Chowning (1980) vividly demonstrated how coordinated frequency modulations would cause computer-generated voices to fuse, whereas miscoordinated modulations would cause the sounds to segregate. (This phenomenon was used in Chowning's 1979 composition, *Phone*.) Bregman and Doehring (1984) experimentally showed that tonal

fusion is enhanced significantly when two tones are modulated by correlated changes in log-frequency. Indeed, the degree of tonal fusion is greater for tones of changing pitch than for tones of static pitch.

In an extensive series of experiments, McAdams (1982, 1984a, 1984b) demonstrated that co-modulations of frequency that preserve the frequency ratios of partials promote tonal fusion. Moreover, McAdams also showed that positively correlated pitch motions that are not precise with respect to log-frequency also tend to contribute to tonal fusion. In other words, tonal fusion is most salient when co-modulation is precise with respect to log-frequency and the frequencies of the two tones are harmonically related. Tonal fusion is next most salient when co-modulation is precise with respect to log-frequency and the frequencies of the two tones are not harmonically related. Finally, tonal fusion is next most salient when co-modulation is positively correlated, but not precise with respect to log-frequency.¹⁰

Apart from the empirical evidence, the pitch co-modulation principle is evident in musical practice as well. In traditional music theory, theorists distinguish two types of positively correlated pitch motion: *similar* motion and *parallel* motion. These types of positively correlated pitch motions might be collectively referred to as *semblant* motions. In Huron (1989a), it was shown that polyphonic composers (not surprisingly) avoid semblant pitch motions—both parallel and similar contrapuntal motions. Moreover, it was shown that parallel pitch motions are avoided more than similar motions. Finally, it was shown that parallel motions are most avoided in the case of intervals that tend most to promote tonal fusion: unisons, octaves, and perfect fifths in particular. Once again, both traditional musical terminology and musical practice are consistent with the perceptual evidence.

Drawing on this research, we might formulate the following principle:

6. Pitch Co-modulation Principle. *The perceptual union of concurrent tones is encouraged when pitch motions are positively correlated. Perceptual fusion is most enhanced when the correlation is precise with respect to log frequency.*

Part II: Derivation of the Rules of Voice-Leading

For the music theorist, the observations made in Part I are highly suggestive about the origin of the traditional rules of voice-leading. It is appropriate, however, to make the relationship between the perceptual principles and the voice-leading rules explicit. In this section, we will attempt to de-

10. It should be noted that psychoacoustic research by Robert Carlyon (Carlyon, 1991, 1992, 1994) suggests that the auditory system may not directly use frequency co-modulation as the mechanism for segregating sources with correlated pitch motions.

rive the voice-leading rules from the six principles described in the preceding section. The derivation given here is best characterized as heuristic rather than formal. The purpose of this exercise is to clarify the logic, to avoid glossing over details, to more easily expose inconsistencies, and to make it easier to see unanticipated repercussions.

In the ensuing derivation, the following abbreviations are used:

- G goal
- A empirical axiom (i.e., an experimentally determined fact)
- C corollary
- D derived musical rule or heuristic
(traditional: commonly stated in music theory sources)
- [D] derived musical rule or heuristic (nontraditional)

In formal logic, an *axiom* is defined as a basic proposition (a statement) that is asserted without proof. However, in this case, each of the axioms presented is supported by the experimental research outlined in Part I. Thus, we might refer to these axioms as “empirical axioms.”

First, we can define the goal of voice-leading:

G1. *The goal of voice-leading is to create two or more concurrent yet perceptually distinct “parts” or “voices.” Good voice-leading optimizes the auditory streaming.*

Two corollaries follow:

C1a. *Effective voice-leading requires clear auditory stream integration within each of the individual parts.*

C1b. *Effective voice-leading requires clear auditory stream segregation between each of the concurrent parts.*

We may begin by introducing our first empirical axiom: the toneness principle:

A1. *Coherent auditory streams are best created using tones that evoke clear auditory images.*

A1a. *Unambiguously pitched complex tones evoke the clearest auditory images.*

From this we derive a nontraditional proposition or rule:

[D1.] *Toneness Rule. Voice-leading should employ tones that evoke strong, unique pitch sensations. This is best achieved using harmonic complex tones.*

[C2.] *Voice-leading is less effective using noises or tones with inharmonic partials.*

In light of these nontraditional rules, a few pertinent observations may be made about musical practice. First, most of the world's music making relies on instruments that produce tones exhibiting harmonic spectra. Instruments that produce inharmonic spectra (e.g., drums) are less likely to be used for voice-leading or melodic purposes. Those percussion instruments that are used melodically (e.g., glockenspiels, marimbas, gamelans, steel drums) are more apt to produce pitched tones. Carilloneurs have long noted that carillons are ill-suited to contrapuntal writing (Price, 1984, p. 310).

Continuing with the toneness principle:

A1b. *Complex tones evoke the strongest sensation of pitch in the region near 300 Hz—between about 80 and 800 Hz.*

From this we derive the rule regarding registral compass:

D2. *Registral Compass Rule. Voice-leading is best practiced in the region between F_2 and G_3 , roughly centered near D_4 .*

Adding the principle of temporal continuity:

A2. *In order to evoke coherent auditory streams, it is best to use continuous rather than intermittent sound sources.*

This leads to a simple nontraditional rule:

[D3.] *Sustained Tones Rule. In general, effective voice-leading is best assured by employing sustained tones in close succession, with few silent gaps or interruptions.*

Adding the minimum masking principle:

A3. *Auditory masking is reduced when partials are evenly spaced with respect to critical bands.*

From this we derive one traditional, and one nontraditional rule of chordal tone spacing:

D4. *Chord Spacing Rule. In general, chordal tones should be spaced with wider intervals between the lower voices.*

[D5.] *Tessitura-Sensitive Spacing Rule. It is more important to have large intervals separating the lower voices in the case of sonorities that are lower in overall pitch.*

Recall that Huron and Sellmer (1992) showed that musical practice is indeed consistent with this latter rule—as predicted by Plomp and Levelt (1965).

Adding the principle of tonal fusion:

A4. *Effective stream segregation is violated when tonal fusion arises.*

The degree to which two concurrent tones fuse varies according to the interval separating them.

A4a. *Tonal fusion is greatest with the interval of a unison.*

From this we derive the traditional injunction against unisons:

D6. Avoid Unisons Rule. *Avoid shared pitches between voices.*

A4b. *Tonal fusion is next strongest at the interval of an octave.*

[D7.] Avoid Octaves Rule. *Avoid the interval of an octave between two concurrent voices.*

A4c. *Tonal fusion is next strongest at the interval of a perfect fifth.*

[D8.] Avoid Perfect Fifths Rule. *Avoid the interval of a perfect fifth between two concurrent voices.*

In general,

[D9.] Avoid Tonal Fusion Rule. *Avoid unisons more than octaves, and octaves more than perfect fifths, and perfect fifths more than other intervals.*

Recall that in Huron (1991b), Bach's polyphonic practice was shown to be consistent with the above nontraditional rules. As simple static intervals, unisons, octaves, and fifths occur less frequently than occurs in a random juxtaposition of parts.

Turning now to the pitch proximity principle:

A5. *The coherence of a stream is maintained by close pitch proximity in successive notes within a voice.*

A4a. *The closest possible proximity occurs when there is no pitch movement.*

Therefore:

D10. Common Tone Rule. *Pitch-classes common to successive sonorities are best retained as a single pitch that remains in the same voice.*

A5b. *The next best stream cohesion arises when pitch movement is near or within van Noorden's "fission boundary" (roughly 2 semitones or less).*

Hence:

D11. Conjunct Movement Rule. *If a voice cannot retain the same pitch, it should preferably move by step.*

C3. Avoid Leaps Rule. *Avoid wide pitch leaps.*

(This last rule is merely a corollary of the Common Tone and Conjunct Movement rules.) When pitch movements exceed the fission boundary, effective voice-leading is best maintained when the temporal coherence boundary is not also exceeded:

A5c. *Wide melodic leaps most threaten auditory stream cohesion when they exceed the temporal coherence boundary.*

From this we derive a nontraditional rule:

[D12.] Leap-Lengthening Rule. *Where wide leaps are unavoidable, use long durations for either one or both of the tones forming the leap.*

Leap-lengthening (consistent with Fitts' law) has been observed in a cross-cultural sample of melodies by Huron and Mondor.

Bregman (1990) has characterized streaming as a *competition* between possible alternative organizations. It is not simply a matter that two successive pitches need to be relatively close together in order to form a stream. The pitches must be closer than other possible pitch-time traces. That is, previous pitches compete for subsequent pitches:

A5d. *Pitches are likely to stream to the nearest previous pitch.*

From this, we can derive those rules of voice-leading that relate to pitch-proximity competition:

D13. Nearest Chordal Tone Rule. *Parts should connect to the nearest chordal tone in the next sonority.*

D14. Part-Crossing Rule. *Avoid the crossing of parts with respect to pitch.*

In a study of part-crossing in 105 works by J. S. Bach, Huron (1991a) showed that of three competing interpretations of the injunction against part-crossing, actual musical practice was most consistent with the principle of pitch proximity.

D15. Pitch Overlapping Rule. *Avoid "overlapped" parts in which a pitch in an ostensibly lower voice is higher than the subsequent pitch in an ostensibly higher voice.*

Once again, failure to obey this rule will conflict with the pitch proximity principle. Each pitch might be likened to a magnet, attracting the nearest subsequent pitch as its partner in the stream.¹¹

Note that a simple general heuristic for maintaining within-voice pitch proximity and avoiding proximity between voices is to place each voice or part in a unique pitch region or tessitura. As long as voices or parts remain in their own pitch “territory,” there is little possibility of part-crossing, overlapping, or other proximity-related problems. There may be purely idiomatic reasons to restrict an instrument or voice to a given range, but there are also good perceptual reasons for such restrictions—provided the music-perceptual goal is to achieve optimum stream segregation. It is not surprising that in traditional harmony, parts are normally referred to by the names of their corresponding tessituras: soprano, alto, and so on.

Adding now the pitch co-modulation principle:

A6a. Effective stream segregation is thwarted when tones move in a positively correlated pitch direction.

[D16.] Semblant Motion Rule. *Avoid similar or parallel pitch motion between concurrent voices.*

A6b. Effective stream segregation is especially thwarted when tones move in precisely positively correlated pitch direction.

[D17.] Parallel Motion Rule. *Avoid parallel motion more than similar motion.*

Huron (1989a) showed that polyphonic composers do indeed tend to avoid positively correlated pitch motions. Moreover, parallel motion tends to be avoided for all interval sizes—not just in the case of tonally fused intervals such as unisons, octaves, and perfect fifths. In addition, parallel motion is avoided more than similar motion.

Derivations From Multiple Principles

Each of the preceding voice-leading rules is deduced from a single perceptual principle. Where two or more principles pertain to the same goal, then transgressing more than one such principle will have more onerous consequences than transgressing only one of the principles. Conversely, the consequences of transgressing a single such principle may be minimized by ensuring complete conformity with the remaining principles—provided all principles pertain to the same goal. The six principles outlined earlier all pertain to the optimization of auditory images and streaming.

11. This analogy is perhaps deceptive because streaming is also known to depend on repetitive and cumulative properties as well as immediate tone successions. See Bregman (1978).

There are other principles that relate to the optimization of auditory streaming, but we will defer deriving the associated rules until Part III. Even in the case of the six stream-related principles identified earlier, 720 (6 factorial) propositions arise from permuting all the principles. We will present only some of these derivations.

If we link the tonal fusion and pitch proximity principles:

A4&A5. The detrimental effect of tonal fusion can be reduced by ensuring close pitch proximity within the parts.

This leads to two nontraditional rules:

[D18.] Oblique Approach to Fused Intervals Rule. *When approaching unisons, octaves, or fifths, it is best to retain the same pitch in one of the voices (i.e., approach by oblique motion).*

[D19.] Avoid Disjunct Approach to Fused Intervals Rule. *If it is not possible to approach unisons, octaves and fifths by retaining the same pitch (oblique motion), step motion should be used.*

A4&A6. Tonal fusion is further encouraged when tonally fused intervals move in a positively correlated direction.

From this we may derive the

[D20.] Avoid Semblant Approach Between Fused Intervals Rule. *Avoid similar pitch motion in which the voices employ unisons, octaves, or perfect fifths. (For example, when both parts ascend beginning an octave apart, and end a fifth apart.)*

A4&A6a. Tonal fusion is most encouraged when tones move in a precisely positively correlated pitch direction at the interval of a unison, octave, or perfect fifth.

From this we derive the well-known prohibition against parallel fifths and octaves:

D21. Parallel Unisons, Octaves, and Fifths Rule. *Avoid parallel unisons, octaves, or fifths.*

Joining the tonal fusion, pitch proximity, and pitch co-modulation principles, we find that:

A4&A5b&A6. When moving in a positively correlated direction toward a tonally fused interval, the detrimental effect of tonal fusion can be alleviated by ensuring proximate pitch motion.

From this we can derive the traditional injunction against exposed or hidden intervals.

D22. Exposed Intervals Rule. *When approaching unisons, octaves, or fifths, by similar motion, at least one of the voices should move by step.*

Note that in traditional treatises on voice-leading most theorists restrict the exposed intervals rule to the case of approaching an octave (fifteenth, etc.)—that is, “exposed octaves.” Only a minority of theorists extend this injunction to approaching perfect fifths (twelfths, etc.) as well. Since octaves engender tonal fusion more easily than perfect fifths, the perceptual principles account for why there would be greater unanimity of musical opinion against exposed octaves than for exposed fifths.

Observe that the principles outlined here are not able to account for the fact that most theorists restrict the exposed intervals rule to voice-leading involving the bass and soprano voices. Huron (1989b) has shown that outer voices are more easily perceived than inner voices; hence one might argue that voice-leading transgressions in outer voices are more noticeable. However, one might just as convincingly argue that since the streaming of the inner voices is more apt to be confounded, it is more important that inner voices conform to the voice-leading principles.

The exposed octaves rule is especially revealing. In essence, the exposed octaves rule says: “If you are going to violate principle A, and if you are going to violate principle B, be certain not to violate principle C as well.” The exposed octaves rule inextricably links three perceptual principles together and logically implies that all three principles share some sort of common goal. Specifically, the exposed octaves rule confirms that the principle of tonal fusion, the pitch proximity principle, and the pitch co-modulation principle all contribute to the achievement of the same goal—namely, the optimization of stream segregation.

Part III: Perceptual Principles and Musical Genres

As in any deductive system, changing one or two axioms will result in a different set of derivations. As in the case of non-Euclidean geometry, it is theoretically possible for composers to devise different voice-leading systems by adding, eliminating, or modifying one or more of the axioms. Changing the axioms will constrain the canon in distinctive ways and hence result in different “musics.”

AUXILIARY PRINCIPLES

In the remainder of this article, we consider four examples of auxiliary perceptual principles. When these auxiliary principles are added to the six core principles, a number of additional propositions arise. As noted in the

introduction, these auxiliary principles in effect constitute voice-leading “options” that will shape the music making in perceptually distinctive ways. At the end of our discussion, we will identify how these auxiliary principles may contribute to the differentiation of musical genres.

The four auxiliary principles discussed in the remainder of this article include (7) onset synchrony, (8) limited density, (9) timbral differentiation, and (10) source location.

7. The Principle of Onset Synchrony

Sounds are more apt to be perceived as components of a single auditory image when they are coordinated in time. That is, concurrent tones are much more apt to be interpreted by the auditory system as constituents of a single complex sound event when the tones are temporally aligned. Because judgments about sounds tend to be made in the first few hundred milliseconds, the most important aspect of temporal coordination is the synchronization of sound *onsets*. Sounds whose onsets are uncoordinated in time are likely to be perceived as distinct or separate events (Bregman & Pinker, 1978). Onset synchronization may be regarded as a special case of amplitude co-modulation. Apart from the crude coordination of tone onsets, the importance of correlated changes of amplitude has been empirically demonstrated for much more subtle deviations in amplitude. For example, Bregman, Abramson, Doehring, and Darwin (1985) have demonstrated that the co-evolution of amplitude envelopes contributes to the perception of tonal fusion.

The concept of synchronization raises the question of what is meant by “at the same time”? Under ideal conditions, it is possible for listeners to distinguish the order of two clicks separated by as little as 20 ms (Hirsh, 1959). However, in more realistic listening conditions, onset differences can be substantially greater and yet retain the impression of a single onset. Initial transients for natural complex tones may be spread over 80 ms or more, even for tones with rapid attack envelopes (Saldanha & Corso, 1964; Strong & Clark, 1967). Many sounds have much more gradual envelopes, and in the case of echoes, more than 100 ms between repetitions may be required in order for successive sounds to be perceived as distinct events (Beranek, 1954).

In the case of the synchronization of note onsets in music, Rasch (1978, 1979, 1988) has collected a wealth of pertinent data. Ensemble performances of nominally concurrent notes show that onsets are typically spread over a range of 30 to 50 ms. Moreover, in experiments with quasi-simultaneous musical tones, Rasch found that the effect of increasing asynchrony is initially to add “transparency” to multitone stimuli—rather than to evoke separate sound images (Rasch, 1988, p. 80). Onset asynchronies have to be comparatively large before separate sound events are perceived.

Although musical notation provides only a crude indication of the temporal aspects of real performances, Rasch's work suggests that synchronously and asynchronously notated events are likely to be perceived as such. Of course tones that are notated as having concurrent onsets are by no means performed precisely together. But in typical performances, such tones are obviously more likely to promote the perception of tonal fusion than tones that are notated as having asynchronous onsets. A difference of a sixteenth-duration at a tempo of 80 quarter notes per minute produces a nominal asynchrony of 187 ms, a time delay that is more than sufficient to elicit the perception of separate events. In short, onsets that are deemed "asynchronous" according to musical notation are unlikely to be perceived as synchronous, whereas onsets that are deemed "synchronous" according to musical notation may well be perceived as being synchronous. Thus musical notation provides a useful (although not infallible) indication of perceived onset synchronies and asynchronies in composed music.

Drawing on this research, we might formulate the following principle:

7. Onset Synchrony Principle. If a composer intends to write music in which the parts have a high degree of perceptual independence, then synchronous note onsets ought to be avoided. Onsets of nominally distinct sounds should be separated by 100 ms or more.

In light of the onset synchrony principle, musical practice is equivocal. Some musics conform to this principle, whereas other musics do not. Rasch (1981) carried out an analysis of vocal works by Praetorius and showed that onset asynchrony between the parts increases as the number of voices is increased. Similarly, Huron (1993a) showed that J. S. Bach minimizes the degree of onset synchrony in his polyphonic works. Figure 11 shows onset synchrony autophases for a random selection of 10 of Bach's 15 two-part keyboard Inventions (BWV's 772-786). A mean onset synchrony function is also plotted (solid line). The values in the horizontal center of the graph indicate the degree of onset synchrony for the work; all other values show the degree of onset synchrony for control conditions in which the parts have been shifted with respect to each other, measure by measure. The dips at zero degrees are indicative of Bach's efforts to avoid synchronous note onsets between the parts.

By contrast, in nonpolyphonic works, composers routinely contradict this principle. Figure 12 compares onset autophases for two different works. The first graph shows an autophase for the three-part Sinfonia No. 1 (BWV 787). The second graph shows an autophase for the four-part hymn "Adeste Fideles." The y-axis values at (0,0) in each x-z plane indicate the degree of onset synchrony for the work. All other values show the degree of onset synchrony for control conditions in which the parts have been systematically shifted with respect to each other, measure by measure. Because there

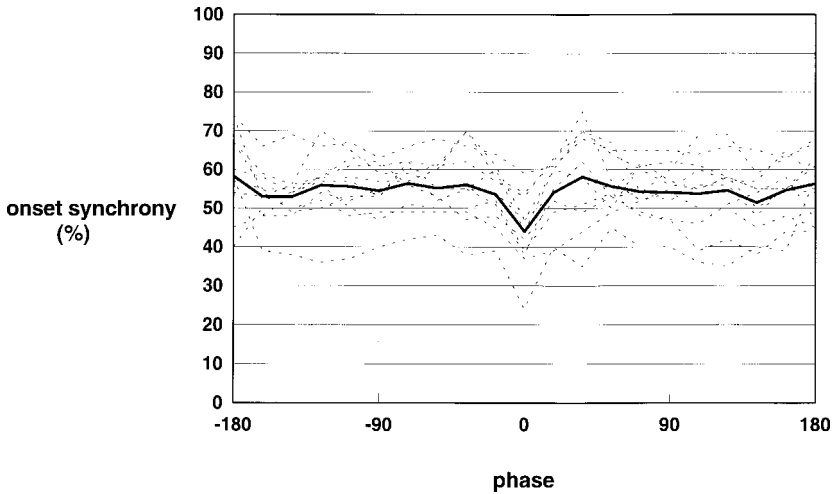


Fig. 11. Onset synchrony autophases for a random selection of 10 of Bach's 15 two-part keyboard Inventions (*BWVs 772–786*); see Huron (1993a). Values plotted at zero degrees indicate the proportion of onset synchrony for the actual works. All other phase values indicate the proportion of onset synchrony for rearranged music—controlling for duration, rhythmic order, and meter. The dips at zero degrees are consistent with the hypothesis that Bach avoids synchronous note onsets between the parts. Solid line plots a mean onset synchrony function for all data.

are more than two parts, the resulting autophases are multidimensional rather than two dimensional. The four-part “Adeste Fideles” has been simplified to three dimensions by phase shifting only two of the three voices with respect to a stationary soprano voice. As can be seen, in the *Sinfonia No. 1* there is a statistically significant dip at (0,0) consistent with the hypothesis that onset synchrony is actively avoided. In “Adeste Fideles” by contrast, there is a significant peak at (0,0) (at the back of the graph) consistent with the hypothesis that onset synchrony is actively sought. The graphs formally demonstrate that any realignment of the parts in Bach's *Sinfonia No. 1* would result in greater onset synchrony, whereas any realignment of the parts in “Adeste Fideles” would result in less onset synchrony.

Musicians recognize this difference as one of musical texture. The first work is polyphonic in texture whereas the second work is homophonic in texture. In a study comparing works of different texture, Huron (1989c) examined 143 works by several composers that were classified a priori according to musical texture. Discriminant analysis¹² was then used in or-

12. Discriminant analysis is a statistical procedure that allows the prediction of group membership based on functions derived from a series of prior classified cases. See Lachenbruch (1975).

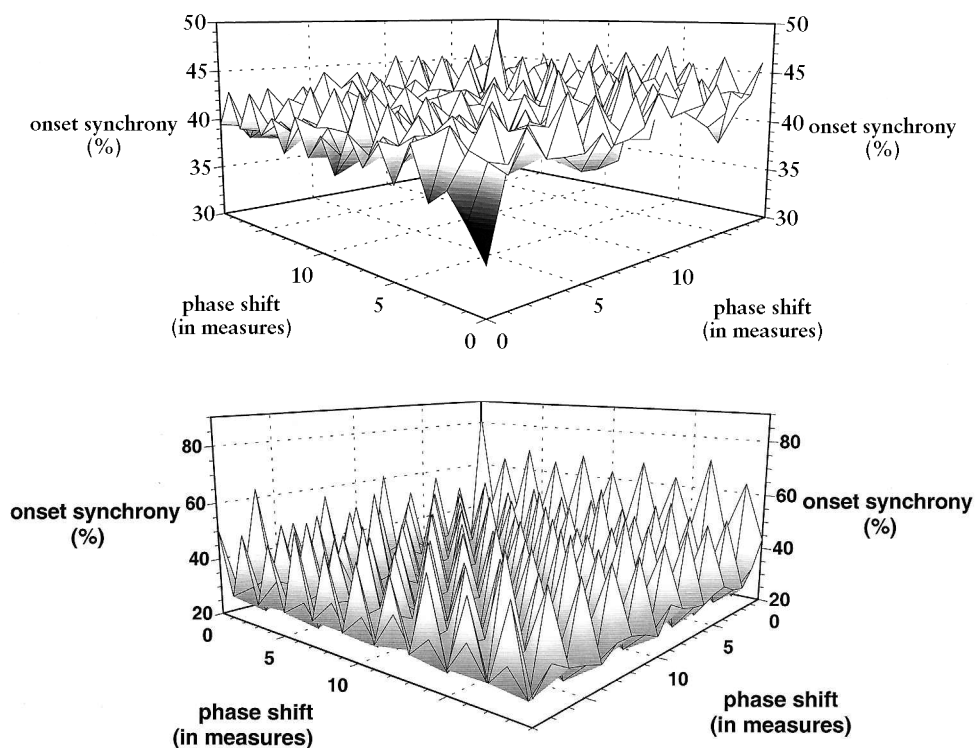


Fig. 12. Three-dimensional onset synchrony autophases comparing exemplar polyphonic and homophonic works. (a) three-part Sinfonia No. 1 (*BWV 787*) by J. S. Bach; (b) four-part hymn “*Adeste Fideles*.” The vertical axis indicates the onset synchrony measured according to “Method 3” described in Huron (1989a, pp. 122–123). The horizontal axes indicate measure-by-measure shifts of two of the voices with respect to the remaining voice(s). Figure 12a shows a marked dip at the origin (front), whereas Figure 12b shows a marked peak at the origin (back). The horizontal axes in the two graphs have been reversed in order to maintain visual clarity. The graphs formally demonstrate that any realignment of the parts in Bach’s Sinfonia No. 1 would result in greater onset synchrony, whereas any realignment of the parts in “*Adeste Fideles*” would result in less onset synchrony.

der to discern what factors distinguish the various types of textures. Of six factors studied, it was found that the most important discriminator between homophonic and polyphonic music is the degree of onset synchrony.

As we have noted, an auxiliary principle (such as the onset synchrony principle) can provide a supplementary axiom to the core principles and allow further voice-leading rules to be derived. Once again, space limitations prevent us from considering all of the interactions. However, consider the joining of the onset synchrony principle with the principal of tonal fusion. In perceptual experiments, Vos (1995) has shown that there is an interaction between onset synchrony and tonal fusion in the formation of auditory streams.

A4&A7. *When approaching a tonally fused interval, the detrimental effect of tonal fusion can be alleviated somewhat by ensuring asynchronous tone onsets.*

From this we can derive the following nontraditional rule:

[D23]. Asynchronous Preparation of Tonal Fusion Rule. *When approaching unisons, octaves, or fifths, avoid synchronous note onsets.*

Figure 13 shows the results from an unpublished study of the interaction between onset synchrony and tonal fusion for polyphonic works (fugues) and homophonic works (chorales) by J. S. Bach. In both homophonic and polyphonic repertoires, dissonant intervals are “prepared” via asynchronous onsets. Imperfect intervals in both repertoires show a tendency to favor synchronous note onsets. However, for perfect intervals, the majority of occurrences are asynchronous in the polyphonic repertoire, whereas the majority are synchronous in the homophonic repertoire. Moreover, for the perfect intervals alone, there is a positive correlation between the degree to which an interval promotes tonal fusion and the likelihood of asynchronous occurrence. In short, the above nontraditional rule is obeyed by Bach—but only in his polyphonic writing.

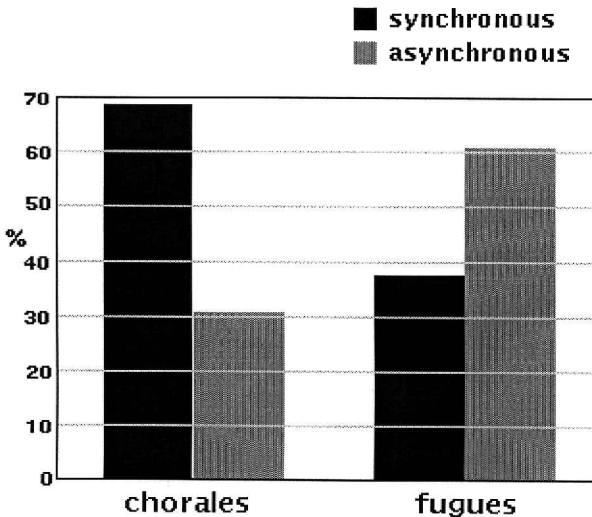


Fig. 13. Results of a comparative study of tonally fused intervals in polyphonic and (predominantly) homophonic repertoires by J. S. Bach. Only perfect harmonic intervals are plotted (i.e., fourths, fifths, octaves, twelfths, etc.). As predicted, in polyphonic textures, most perfect intervals are approached with asynchronous onsets (one note sounding before the onset of the second note of the interval). In the more homophonic chorale repertoire, most perfect intervals are formed synchronously; that is, both notes tend to begin sounding at the same moment. By contrast, the two repertoires show no differences in their approach to imperfect and dissonant intervals. From Huron (1997).

Whence Homophony?

The existence of homophonic music raises an interesting perceptual puzzle. On the one hand, most homophonic writing obeys all of the traditional voice-leading rules outlined in Part I of this article. (Indeed, most four-part harmony is written within a homophonic texture.) This suggests that homophonic voice-leading is motivated (at least in part) by the goal of stream segregation—that is, the creation of perceptually independent voices. However, because onset asynchrony also contributes to the segregation of auditory streams, why would homophonic music not also follow this principle? In short, why does homophonic music exist? Why isn't all multipart music polyphonic in texture?

The preference for synchronous onsets may originate in any of a number of music-related goals, including social, historical, stylistic, aesthetic, formal, perceptual, emotional, pedagogical, or other goals that the composer may be pursuing. If we wish to posit a *perceptual* reason to account for this apparent anomaly, we must assume the existence of some additional perceptual goal that has a higher priority than the goal of stream segregation. More precisely, this hypothetical perceptual goal must relate to some aspect of *temporal* organization in music. Two plausible goals come to mind.

In the first instance, the texts in vocal music are better understood when all voices utter the same syllables concurrently. Accordingly, we might predict that vocal music is less likely to be polyphonic than purely instrumental music. Moreover, when vocal music is truly polyphonic, it may be better to use melismatic writing and reserve syllable onsets for occasional moments of coincident onsets between several parts. In short, homophonic music may have arisen from the concurrent pursuit of two perceptual goals: the optimization of auditory streaming, and the preservation of the intelligibility of the text. Given these two goals, it would be appropriate to abandon the pursuit of onset asynchrony in deference to the second goal.

Another plausible account of the origin of homophony may relate to rhythmic musical goals. Although virtually all polyphonic music is composed within a metric context, the rhythmic effect of polyphony is typically that of an ongoing “braid” of rhythmic figures. The rhythmic strength associated with marches and dances (pavans, giges, minuets, etc.) cannot be achieved without some sort of rhythmic uniformity. Bach's three-part Sinfonia No. 1 referred to in connection with Figure 12 appears to have little of the rhythmic drive evident in “Adeste Fideles.” In short, homophonic music may have arisen from the concurrent goals of the optimization of auditory streaming, and the goal of rhythmic uniformity. As in the case of the “text-setting” hypothesis, this hypothesis generates a prediction—namely, that works (or musical forms) that are more rhythmic in character (e.g., cakewalks, sarabandes, waltzes) will be less polyphonic than works following less rhythmic forms.

8. The Principle of Limited Density

Listeners do not have an unbounded capacity to track multiple concurrent lines of sound. Huron (1989b) found that as the number of concurrent voices in a polyphonic texture increases, expert listeners are both slower to recognize the addition of new voices and more prone to underestimate the number of voices present. Figure 14 plots experimental data from five expert musicians. The solid columns show mean estimation errors for polyphonic textures of various textural densities; the shaded columns indicate mean error rates for single-voice entries. For polyphonic textures employing relatively homogeneous timbres, the accuracy of identifying the number of concurrent voices drops markedly at the point where a three-voice texture is augmented to four voices. Beyond three voices, tracking confusions are commonplace. In a study of isolated chords, Parncutt (1993) found similar results. When asked to count the number of tones in chords constructed using tones with octave-spaced partials (Shepard tones), Parncutt found that listeners make significant errors (underestimation) once the number of chromas exceeds three.

Limitations in the perception of auditory numerosity may be symptomatic of a broader perceptual limitation, because the results of these studies parallel the results of similar studies in vision. A variety of counting and estimation tasks show a marked increase in confusion when the number of

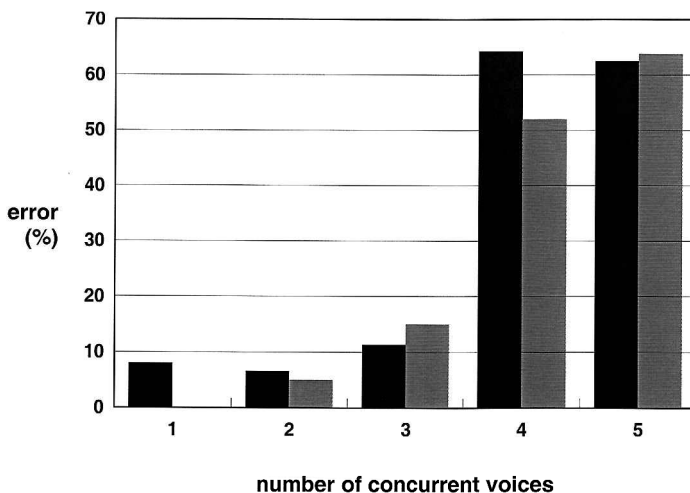


Fig. 14. Voice-tracking errors while listening to polyphonic music. Solid columns: mean estimation errors for textural density (no. of polyphonic voices); data for 130 trials from five expert musician subjects. Shaded columns: unrecognized single-voice entries in polyphonic listening; data for 263 trials from five expert musician subjects (Huron, 1989b). The data show that when listening to polyphonic textures that use relatively homogeneous timbres, tracking confusions are common when more than three voices are present.

visual items exceeds three (Estes & Combes, 1966; Gelman & Tucker, 1975; Schaeffer, Eggleston, & Scott, 1974). Descoedres (1921) characterized this as the *un, deux, trois, beaucoup* phenomenon. In a study of infant perceptions of numerosity, Strauss and Curtis (1981) found evidence of this phenomenon in infants less than a year old. Using a dishabituation paradigm, Strauss and Curtis showed that infants are readily able to discriminate two from three visual items. However, performance degrades when discriminating three from four items and reaches chance performance when discriminating four from five items. The work of Strauss and Curtis is especially significant because preverbal infants can be expected to possess no explicit knowledge of counting. This implies that the perceptual confusion arising for visual and auditory fields containing more than three items is a low-level constraint that is not mediated by cognitive skills in counting.

These results are consistent with reports of musical experience offered by musicians themselves. Mursell (1937) reported a conversation with Edwin Stringham in which Stringham noted that “no human being, however talented and well trained, can hear and clearly follow more than three simultaneous lines of polyphony.” A similar claim was made by composer Paul Hindemith (1944).

Drawing on this research, we might formulate the following principle:

8. Principle of Limited Density. *If a composer intends to write music in which independent parts are easily distinguished, then the number of concurrent voices or parts ought to be kept to three or fewer.*

Of course most harmony writing consists of *four* or more parts. Indeed, most harmony texts are based exclusively on four-part writing. So once again, the question that arises regarding this principle is why it is rarely followed.

As a start, we must first recognize that a surprising amount of music actually conforms to the above principle of limited density. Consider, for example, the density of polyphonic writing in the music of J. S. Bach. Figure 15 (Huron, 1989a) shows the estimated number of auditory streams versus the nominal number of voices or parts for a large body of polyphonic music by Bach. Estimates of the number of auditory streams are calculated according to a method described and tested experimentally by Huron (1989a). The dotted line indicates a relation of unity slope. If an N -part work truly maintains a mean textural density of N auditory streams, then the plotted value for the work would be expected to lie near the unity slope (dotted line). In order for a value to lie above the dotted line, the work must contain parts that manifest a considerable degree of pseudopolyphonic activity and must also refrain from much voice retirement (idle parts). Conversely, in order for a value to lie below the dotted line, the work must contain significant periods of voice retirement and refrain from

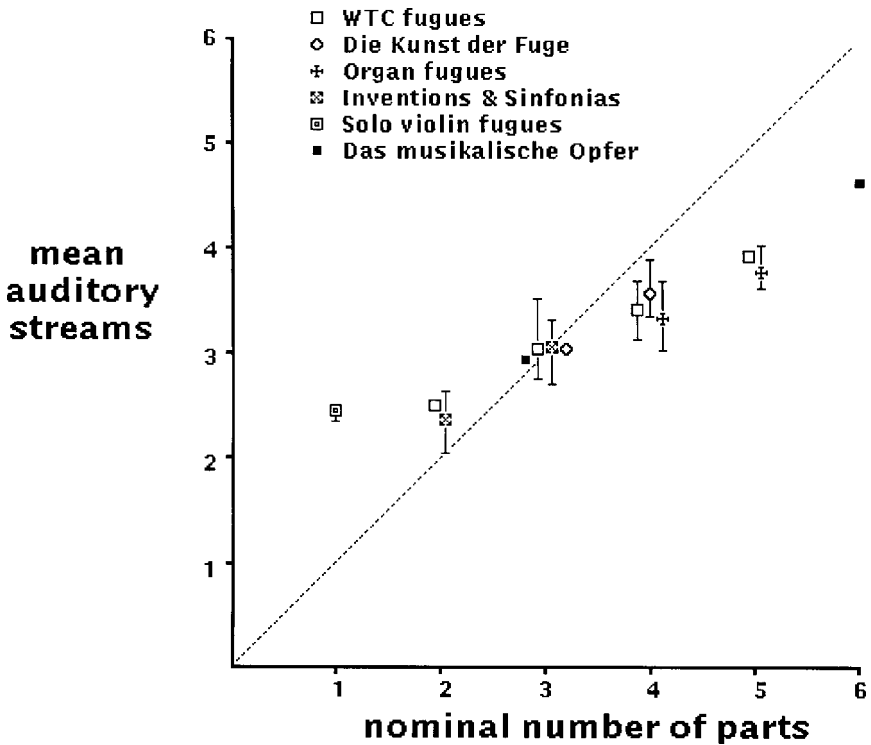


Fig. 15. Relationship between nominal number of voices or parts and mean number of estimated auditory streams in 108 polyphonic works by J. S. Bach. Mean auditory streams for each work calculated using the stream-latency method described and tested in Huron (1989a, chap. 14). Bars indicate data ranges; plotted points indicate mean values for each repertoire. Dotted line indicates a relation of unity slope, where an N -voice polyphonic work maintains an average of N auditory streams. The graph shows evidence consistent with a preferred textural density of three auditory streams. See also Huron and Fantini (1989).

pseudo-polyphonic writing. Figure 15 shows that as the number of nominal voices increases, Bach gradually changes his compositional strategy. For works with just two parts, Bach endeavors to keep the parts active (few rests of short duration) and to boost the textural density through pseudo-polyphonic writing. For works with four or more nominal voices, Bach reverses this strategy. He avoids writing pseudo-polyphonic lines and retires voices from the texture for longer periods of time. Figure 15 reveals a change of strategy consistent with a preferred textural density of three auditory streams. In light of the extant research, it would appear that Bach tends to maximize the number of auditory streams, while simultaneously endeavoring to avoid exceeding the listener's ability to track the concurrent parts.

As for works that regularly use four or more concurrent parts, once again, the notion of musical genre may be helpful in understanding diver-

gent practices. In contrast to Bach's polyphonic practice, musical works consisting of four or more concurrent parts are more likely to be perceived as homophonic in texture. By employing such large numbers of concurrent parts, homophonic textures may tend to obscure the component tones and encourage synthetic perceptions such as chords.

9. The Principle of Timbral Differentiation

The influence of timbre (spectral content) on the formation of auditory streams was demonstrated by van Noorden (1975) and Wessel (1979), and it was tested experimentally by McAdams and Bregman (1979), Bregman, Liao, and Levitan (1990), Hartmann and Johnson (1991) and Gregory (1994). Wessel created stimuli consisting of a repeated sequence of three rising pitches (see Figure 16). Successive tones were generated so that every second tone was synthesized by using one of two contrasting timbres. Notice that although the pitch sequence is rising, each of the "timbre sequences" produces a descending sequence.

If timbre-based streaming takes precedence over pitch-based streaming, then listeners ought to hear two distinct descending pitch sequences rather than a single ascending pitch sequence. Wessel found that as the speed of repetition is increased, there is a greater tendency for the timbre differences to cause listeners to hear descending sequences. Also, Wessel found that the effect is most pronounced when the two timbres used are most contrasting (as, for example, when their spectral centroids are far apart). Iverson (1992) has shown that timbral cues that contribute most to streaming are highly correlated with similarity judgments. That is, listeners are most likely to link together sonic events that share similar tone colors; conversely, listeners tend to segregate sonic events whose tone colors are quite different.

In the case of music perception, Erickson (1975) has identified a number of ways in which timbre is used to distinguish musical elements. Rössler (1952) carried out an extensive analysis of the influence of organ registration on the perceptual independence of polyphonic and pseudo-polyphonic lines. Rössler's results underscore the role of timbre in the fission and fusion of auditory streams.

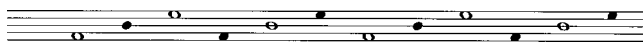


Fig. 16. Schematic illustration of Wessel's illusion (Wessel, 1979). A sequence of three rising pitches is constructed by using two contrasting timbres (marked with open and closed noteheads). As the tempo of presentation increases, the percept of a rising pitch figure is superseded by two independent streams of descending pitch—each stream distinguished by a unique timbre.

Drawing on this research, we might formulate the following principle:

9. *Timbral Differentiation Principle. If a composer intends to write music in which the parts have a high degree of perceptual independence, then each part should maintain a unique timbral character.*

Again, the question that immediately arises regarding this principle is, why do composers routinely ignore it? Polyphonic vocal and keyboard works, string quartets, and brass ensembles maintain remarkably homogeneous timbres. Of course, some types of music do seem to be consistent with this principle. Perhaps the most common Western genre that employs heterogeneous instrumentation is music for woodwind quintet. Many other small ensembles make use of heterogeneous instrumentation. Nevertheless, it is relatively rare for tonal composers to use different timbres for each of the various voices. A number of reasons may account for this. In the case of keyboard works, there are basic mechanical restrictions that make it difficult to assign a unique timbre to each voice. Exceptions occur in music for dual-manual harpsichord and pipe organ. In the case of organ music, a common Baroque genre is the *trio sonata* in which two treble voices are assigned to independent manuals and a third bass voice is assigned to the pedal division. In this case it is common to differentiate polyphonic voices timbrally by using different registrations—indeed, this is the common performance practice. The two-manual harpsichord can also provide contrasting registrations, although only two voices can be distinguished in this manner.

The case of vocal ensembles is more unusual. There are some differences in vocal production, and hence there are modest differences in vocal character for bass, tenor, alto, and soprano singers. However, choral conductors normally do not aim for a clear difference in timbre between the various voices. On the contrary, most choral conductors attempt to shape the sound of the chorus so that a more homogeneous or integrated timbre is produced. Nevertheless, it is possible to imagine a musical culture in which such vocal differences were highly exaggerated. Why isn't this commonplace?

We may entertain a variety of scenarios that might account for the avoidance of timbral differentiation in much music. Some of these scenarios relate to idiomatic constraints in performance. For example, much Western polyphonic music is written for vocal ensemble or keyboard. On the keyboard, we have noted that simple mechanical problems make it difficult to assign different timbres to each voice. In ensemble music, recruiting the necessary assortment of instrumentalists might have raised practical difficulties. It might be easier to recruit two violinists than to recruit a violinist and an oboe player. Such practical problems might have forced composers

to abandon timbral differentiation as a means of enhancing the perceptual independence of the parts.

Once again, if we wish to pursue a perceptual approach to explaining this apparent anomaly, we must posit the existence of some additional perceptual goal that has a higher priority than that of auditory streaming. One plausible goal is that of *balance*—that is, the relative salience or noticeability of the various parts. As long as the parts are performed on similar instruments or are restricted to human voices, no single part is apt to be louder or more noticeable than another. However, once heterogeneous timbral resources are adopted, a multitude of differences are unleashed. One or another timbre may be more commanding of the listener's attention and thus antithetical to maintaining an equivalent status in the auditory scene. A French horn, for example, is no match for an oboe. Provided both instruments are equally busy, the oboe will always tend to dominate the listener's attention. Similarly, a human voice will nearly always be more salient than a musical instrument. Notice that timbral differentiation seems to be preferred in situations where the composer wishes to draw attention to a particular line of sound against a background of other sounds. That is, timbral differentiation may be the primary means for choreographing foreground/background effects.

Another plausible (and related) goal is that of achieving a homogeneous blend of sound. Sandell (1991b) has studied a number of perceptual correlates for orchestral "blend." Composers tend to prefer those instrumental combinations that show a high degree of blend. It may be that homogeneous instrumentation merely provides the easiest way to achieve a blended sound. Although heterogeneous instrumentation might enhance the perceptual segregation of the voices, it would seem that tonal composers have typically found the overall timbral result less preferable to a uniform ensemble.

10. The Source Location Principle

Even in reverberant environments, listeners are adept at identifying the physical location from which a sound originates. It is known, for example, that reflected sound energy is suppressed by the auditory system so that listeners make localization judgments on the basis of the initial (unreflected) sound (Haas, 1951). Although we have only two ears, listeners can distinguish anterior from posterior sources, as well as elevation. Even when listening with just a single ear, humans show significant localization skills. Divenyi and Oliver (1989) have shown that spatial location has a strong effect on streaming (although it has relatively little effect on tonal fusion). Since one of the best generalizations that can be made about independent sound sources is that they normally occupy distinct positions in space, one

would naturally expect location to provide strong streaming cues (Huron, 1991c, p. 78).

Drawing on this research, we might formulate the following principle:

10. Source Location Principle. *If a composer intends to write music in which the parts have a high degree of perceptual independence, then it is helpful to separate in space the various sound sources for each part.*

The applications of this principle in music are largely obvious. In the case of antiphonal writing, sound sources may be separated by 50 m or more. Conversely, when two or more instruments combine to play the same part, they are typically seated in close physical proximity—as in the second violin section in an orchestra.

Other spatial effects are more subtle. In the case of barbershop quartet singing, common performance practice places the singers in very close proximity, with the singers' heads positioned close to a single location in space (Martin, 1965; Root, 1980). This implies a goal of greater perceptual homogeneity—more precisely, the avoidance of stream segregation.

A number of musical works in the late twentieth century have consciously manipulated spatial parameters. In the case of instrumental works, spatial effects are often associated with unorthodox seating arrangements. In extreme cases, such as Iannis Xenakis' *Terretektorh*, the instrumentalists are dispersed throughout the audience. In general, musicians have tended to avoid such wide separations because listeners often find themselves unable to hear much beyond the musician sitting next to them. In *Terretektorh* however, this perceptual goal may be less important than a sociological goal of attempting to break down the traditional dichotomy between audience and performers.

The traditional reluctance of musicians to use dispersed locations for sounds may be readily explained by the ensuing problems of acoustic variance. As sound sources become more widely separated, the musical effect for different audience members becomes more variable: what listeners hear depends primarily on where they sit. Conversely, when all the sound sources are in close physical proximity, each individual source is placed on a more equal footing. As in the case of timbral differentiation, the traditional reluctance to employ spatial differentiation may arise because of problems in maintaining ensemble balance.

In the case of electroacoustic music, spatial manipulation may prove more effective. In the first instance, much electroacoustic music is experienced via stereophonic recordings in personal listening settings—hence reducing the effect of listener location. In addition, spatial cues may be dynamic (trajectories) rather than static (fixed location). By using continually moving virtual sound sources, it may be possible to maintain a relatively homogeneous texture without inadvertent foreground/background effects.

AUXILIARY PRINCIPLES AND MUSICAL TEXTURES

In Part III of this article, we have argued that some perceptual principles are auxiliary to the core voice-leading principles. It has been argued that these principles are treated as compositional “options” and are chosen according to whether or not the composer wishes to pursue additional perceptual goals. If this hypothesis is correct, then measures derived from such auxiliary principles ought to provide useful discriminators for various types of music. For example, the degree of onset synchrony may be measured according to a method described in Huron (1989a), and this measure ought to discriminate between those musics that do, or do not, conform to the principle of onset synchrony.

Figure 17 reproduces a simple two-dimensional “texture space” described by Huron (1989c). A bounded two-dimensional space can be constructed

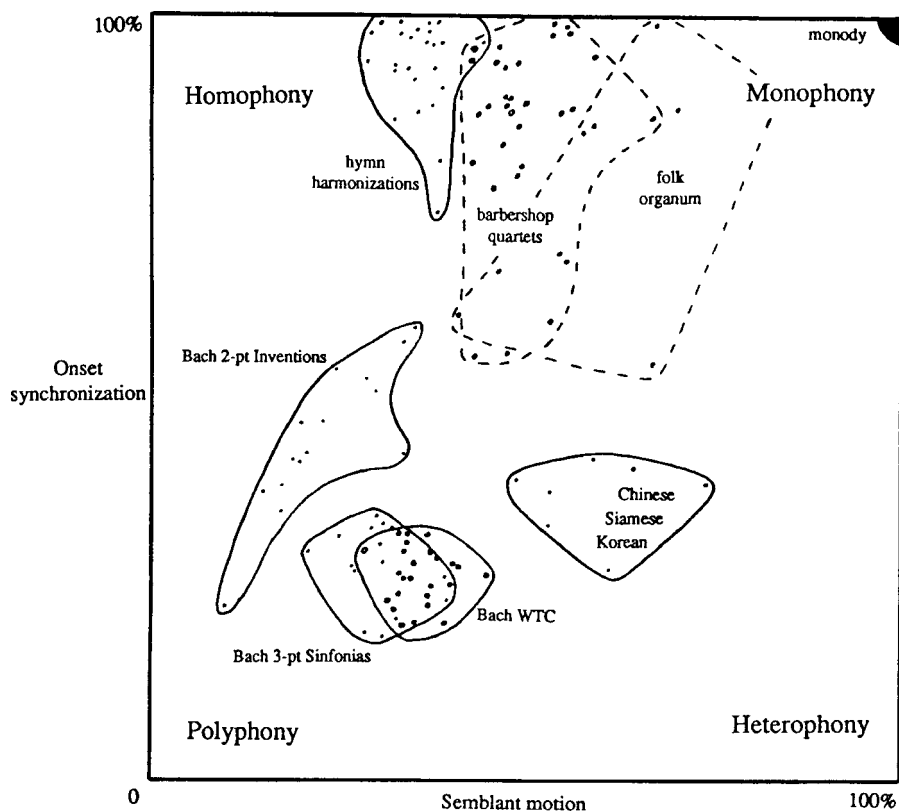


Fig. 17. Texture space reproduced from Huron (1989c). Works are characterized according to two properties: onset synchronization (the degree to which notes or events sound at the same time) and semblant motion (the degree to which concurrent pitch motions are positively correlated). Individual works are plotted as points; works within a single repertoire are identified by closed figures.

using semblant contrapuntal motion and onset synchrony as the x and y axes. By definition, a single monophonic line will be exactly synchronized with itself in terms of onsets and will move in precise pitch motion with itself. As a result, all monophonic works are necessarily plotted in the upper right hand corner. Hymns reside in the upper mid-left area. Polyphonic works by J. S. Bach sit in the lower left area, whereas a small sample of non-Western works are further to the right. The resulting space appears to divide into quadrants representing three of the four principle types of musical textures: monophony, homophony, and polyphony. (A sample of heterophonic music occupies a region closer to the center of the space). Moreover, the conceptual utility of this texture space is evident in its placement of hybrid textures, such as Kentucky folk organum and barbershop quartets. Barbershop quartets are plotted in a region part-way between homophony and monophony—reflecting the greater cohesiveness of close harmony writing compared to hymnody.

Figure 18 shows another view of similar data selected from a large musical data base. Unlike the hand-drawn boundaries plotted in Figure 17, the distributions in Figure 18 arise from the works themselves. In Figure 18, each of 173 works was characterized according to the degree of semblant motion and onset synchronization. From the ensuing discrete data set of x -

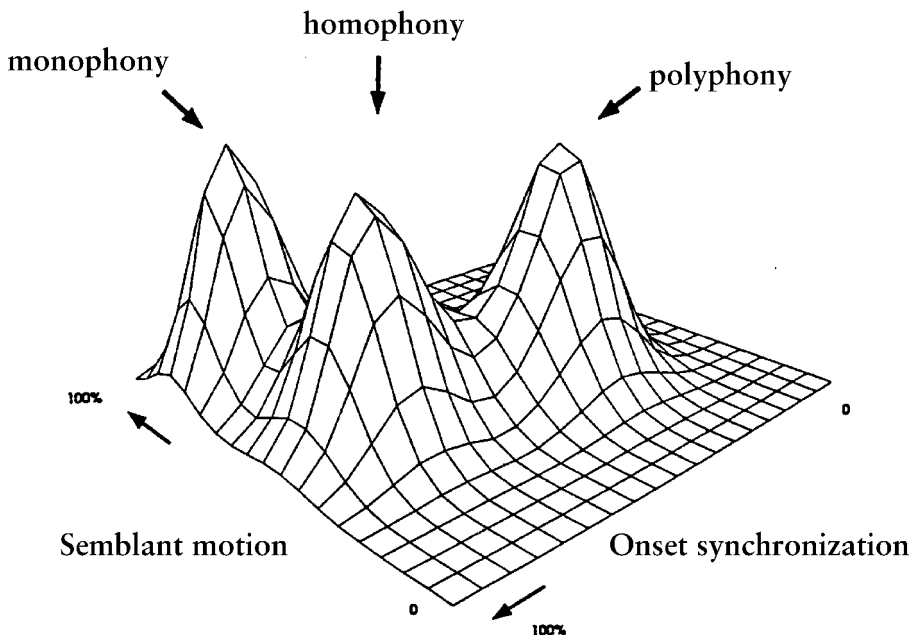


Fig. 18. Distribution of musical textures from a sample of 173 works; generated using Parzen windows method. (The axes have been rearranged relative to Figure 18 in order to maintain visual clarity.)

y points a Parzen windows technique (Duda & Hart, 1973) has been used to estimate the actual distribution shapes. As can be seen, three classic genres—monophony, polyphony, and homophony—are clearly evident. (Heterophonic works were not sufficiently represented in the sample for them to appear using this technique.)

By itself, this “texture space” does not prove the hypothesis that some perceptual principles act to define various musical genres. This texture space is merely an existence proof that, *a priori*, genres can be distinguished by 2 of the 10 principles discussed here. A complete explication of perceptually based musical textures would require a large-scale multivariate analysis that is beyond the scope of this article.

Conclusion

In this article, we have shown that perceptual principles can be used to account for a number of aspects of musical organization, at least with respect to Western music. In the first instance, it was shown that six empirically established perceptual principles are able to account for the majority of traditional rules in Western voice-leading. Moreover, the power of these perceptual principles was evident in the fact that nontraditional rules derived from the principles were also found to be consistent with actual musical practice.

As in a formal system, it is possible to eliminate, modify, or add different perceptual principles—in a manner analogous to axioms in formal logic. The addition, elimination, or modification of different perceptual principles leads to different sonic textures and so can be used to define or distinguish various musical genres.

Of course, the account of voice-leading given here remains incomplete. Unresolved issues include a variety of questions. First, why do theorists typically restrict the injunction against exposed octaves to outer voices only? Because inner voices are more difficult to trace perceptually, one might have thought that inner voices would be more prone to stream integration arising from exposed octaves. Second, little is known about the influence of perceptual schemas in the formation of auditory streams. Bregman (1990) has assembled evidence illustrating the existence and influence of learned schemas in auditory scene analysis. It is possible that motivic and rhythmic features of musical organization facilitate stream segregation, but no research has yet addressed this issue. For example, is it the case that imitative polyphony facilitates or inhibits stream segregation compared with nonimitative polyphony?

Third, the perceptual scenario outlined here does not account for the sense of direction (“leading”) that attends musical pitch successions. For

example, the above account implies that voice-leading should be equally effective when a work is played backward rather than forward.¹³ However, embellishments (e.g., suspensions, appoggiaturas) exhibit systematic asymmetries, such as the penchant for resolution by step, but approach by either step or leap. A plausible cognitive answer may be found in the anchoring of unstable tones to stable ones, as demonstrated by Krumhansl (1979) and investigated more thoroughly by Bharucha (1984).

In addition to these questions, several untested predictions arise from the preceding account of voice-leading. The minimum masking principle would predict that chordal-tone spacing ought to be influenced by spectral content. Hence, the spacing of chords ought to differ in music that uses differing instrumentation. Second, we do not yet understand the relative importance of each of the various factors known to influence auditory streaming. When such empirical knowledge becomes available, then we ought to be able to measure the extent to which musical practice conforms to the relative perceptual weights associated with each principle.

An important question is whether the principles discussed in this article are universal aspects of human auditory perception or whether they arise from culture-specific conditioning. Questions of nature and nurture are important in music because they have implications for how we view music and what we can expect of music's creative future. If perceptual principles—such as those described in this article—are primarily learned, then it suggests that human perceptual capacities have a remarkable fluidity and adaptability. Musicians might take heart in such knowledge, because it implies that there is great potential for new creative directions in musical culture. Conversely, if such principles are found to be physiologically determined, then it suggests that human perceptual capacities are constrained by fixed biological limits. Musicians might be disheartened by such knowledge, because it implies that musical cultures are destined to develop within a finite region of musical possibilities.

Several points should be kept in mind regarding nature/nurture questions. First, beliefs are always likely to be more strongly held than is warranted by the available empirical research. Second, both extremes of belief have possible negative repercussions. Those who believe that perceptual mechanisms are largely fixed and immutable are in danger of underestimating human capabilities and therefore of impeding the development of musical culture. Conversely, those who believe that perceptual mechanisms are highly adaptable are in danger of overestimating human adaptability and therefore of encouraging activities that are irrelevant to human needs. Although questions concerning the adaptability of human perception may invite opinions, these questions more properly invite further empirical investigation.

13. Note onsets and note offsets will be reversed in backwards renditions, and this may affect onset synchronies.

Of course, even if the “rules of voice-leading” were shown to be somehow immutable, their musical relevance depends entirely on the acceptance of the underlying perceptual goals. In short, the composer must deem the goals discussed here worth pursuing. If musicians see no need to pursue these goals, then the principles described here are musically irrelevant. Even when such goals are deemed worthy of pursuit, they must be balanced against other possible compositional goals—such as social, historical, stylistic, aesthetic, or formal goals—or other cognitive or perceptual goals deemed to have a higher priority.

Aesthetic Origins

Given the plethora of possible compositional goals, we might rightly ask why so much Western music appears to be consistent with the perceptual goal discussed here. Why would so many composers endeavor to clarify the auditory streaming of the individual parts in a musical texture?

Bregman (1990) has advanced the hypothesis that music makes use of preexisting auditory mechanisms for parsing the environment. A useful parallel can be drawn to animated cartoons. Although the cartoonist is free to draw anything at all, there are advantages to drawing objects and motions that resemble real-world visual scenes. Viewers will readily apply preexisting visual mechanisms in making sense of fictitious images, even when the objects of manipulation are nonrepresentational. According to Bregman, composers similarly create “auditory fictions” populated by plausible (although largely fictional) auditory objects and events. Music listeners may be predisposed to make use of their existing auditory expertise related to the parsing of auditory scenes.

Absent from Bregman’s account is an aesthetic discussion of why this might be enjoyable for listeners. Why do many listeners take pleasure in music constructed according to the rules of voice-leading?

Like all sensory systems, the auditory system arose as an evolutionary adaptation whose function is to provide listeners with relevant information about the external world. Sensory systems are not infallible in carrying out this task, however. Sensory information can often be incomplete or ambiguous, and deciphering the world entails a certain amount of mental effort. Particularly when a scene is complicated, there is good reason to suppose that the limbic system rewards the brain when perceptual systems successfully “parse” the available information. By “success,” we do not mean that the mental representation constructed by perception accurately reflects the real world—since there is no independent way of judging such accuracy. Rather, by “success” we mean that the mental representation is

coherent, self-consistent, and congruent with information from other sensory systems and schematic expectations.

A number of phenomenal experiences accord with the view that pleasure is related to perceptual success. Watching an out-of-focus film can engender considerable annoyance, followed by a sense of relief when the focus is finally adjusted. A parallel auditory experience is evident when tuning a radio to eliminate noise or static. Switching from monophonic to stereophonic reproduction both improves the spatial resolution of the sound sources and simultaneously evokes pleasure.

An apparent problem with this view arises from composers' penchant to create music containing several concurrent parts. If pleasure is evoked by successful parsing of the auditory scene, then why wouldn't musicians create solely monophonic music? A plausible answer to this question is that the amount of pleasure evoked may be proportional to the perceptual difficulty posed by the scene. Consider, by way of illustration, the experience of random-dot stereograms. These are three-dimensional images that are coded in a superficially confusing array of random dots. Seeing the three-dimensional image can involve considerable mental effort, but if the image is successfully apprehended, there ensues a rush of delight for the viewer. Similarly, it is possible that some multipart music is organized to challenge the listener's auditory parsing capacities.

This interpretation carries several implications. It suggests that early Renaissance polyphonists discovered an aesthetic effect akin to random-dot stereograms: by challenging the listener's auditory parsing abilities, the potential for a pleasing effect could be heightened. However, this heightened pleasure would be possible only if listeners could indeed successfully parse the more complex auditory scenes. Having increased the perceptual challenge, composers would need to take care in providing adequate streaming cues. Following the rules of voice-leading and limiting the density of parts (as we have seen in Bach) might be essential aids for listeners. A further implication is that at least some of the listeners who claim to dislike polyphonic music may fail to "get it" in the same way that some viewers fail to "get" random-dot stereograms.

This interpretation also has implications for how we understand sensory dissonance. Earlier, we interpreted the congruence between chordal-tone spacing and the cochlear map as consistent with minimizing masking. However, the results are equally consistent with the goal of minimizing sensory dissonance. Since both sensory dissonance and auditory masking are linked to the critical band, it may be that the phenomena are two sides of the same coin. The central characteristic of dissonance is the negatively valenced phenomenal experience of listeners—akin to annoyance. Why would listeners find spectral components within a critical band annoying? Perhaps

because the auditory system recognizes the presence of masking, with the implication that any seemingly successful parsing of the auditory scene may be deceptive or faulty.

Note that the preceding interpretation of the aesthetic origins of voice-leading should in no way be construed as evidence for the superiority of polyphonic music over other types of music. In the first instance, different genres might manifest different perceptual goals that evoke aesthetic pleasure in other ways. Nor should we expect pleasure to be limited only to perceptual phenomena. As we have already emphasized, the construction of a musical work may be influenced by innumerable goals, from social and cultural goals to financial and idiomatic goals. Finally, it is evident from Figure 17, that many “areas” of possible musical organization remain to be explored. The identification of perceptual mechanisms need not hamstring musical creativity. On the contrary, it may be that the overt identification of the operative perceptual principles may spur creative endeavors to generate unprecedented musical genres.^{14,15}

References

- Aldwell, E., & Schachter, C. (1989). *Harmony and Voice-Leading*. 2nd ed. Orlando, FL: Harcourt, Brace, Jovanovich.
- ANSI. (1973). *Psychoacoustical Terminology*. New York: American National Standards Institute.
- Attneave, F., & Olson, R. K. (1971). Pitch as a medium: A new approach to psychophysical scaling. *American Journal of Psychology*, 84, 147–166.
- von Békésy, G. (1943/1949). Über die Resonanzkurve und die Abklingzeit der verschiedenen Stellen der Schneckentrennwand. *Akustische Zeitschrift*, 8, 66–76; translated as: On the resonance curve and the decay period at various points on the cochlear partition. *Journal of the Acoustical Society of America*, 21, 245–254.
- von Békésy, G. (1960). *Experiments in hearing*. New York: McGraw-Hill.
- Benjamin, T. (1986). *Counterpoint in the style of J. S. Bach*. New York: Schirmer Books.
- Beranek, L. L. (1954). *Acoustics*. New York: McGraw-Hill. (Reprinted by the Acoustical Society of America, 1986).
- Berardi, A. (1681). *Documenti armonici*. Bologna, Italy: Giacomo Monti.

14. Several scholars have made invaluable comments on earlier drafts of this article. Thanks to Bret Aarden, Jamshed Bharucha, Albert Bregman, Helen Brown, David Butler, Pierre Divenyi (Divenyi, 1993), Paul von Hippel, Adrian Houstma, Roger Kendall, Mark Lochstampor, Steve McAdams, Eugene Narmour, Richard Parncutt, Jean-Claude Risset, Peter Schubert, Mark Schmuckler, David Temperley, James Tenney, and David Wessel.

Special thanks to Gregory Sandell for providing spectral analyses of orchestral tones, to Jasba Simpson for programming the Parzen windows method, and to Johan Sundberg for drawing my attention to the work of Rössler (1952).

15. This article was originally submitted in 1993; a first round of revisions was undertaken while the author was a Visiting Scholar at the Center for Computer Assisted Research in the Humanities, Stanford University. Earlier versions of this article were presented at the Second Conference on Science and Music, City University, London (Huron, 1987), at the Second International Conference on Music Perception and Cognition, Los Angeles, February 1992, and at the Acoustical Society of America 125th Conference, Ottawa (see Huron, 1993c).

- Bharucha, J. J. (1984). Anchoring effects in music: The resolution of dissonance. *Cognitive Psychology*, 16, 485–518.
- Bharucha, J. J. (1992). *The emergence of auditory and musical cognition from neural nets exposed to environmental constraints*. Paper presented at the Second International Conference on Music Perception and Cognition, Los Angeles.
- Bozzi, P., & Vicario, G. (1960). Due fattori di unificazione fra note musicali: la vicinanza temporale e la vicinanza tonale. *Rivista di psicologia*, 54(4), 253–258.
- Bregman, A. S. (1978). Auditory streaming is cumulative. *Journal of Experimental Psychology: Human Perception and Performance*, 4, 380–387.
- Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.
- Bregman, A. S., Abramson, J., Doehring, P., & Darwin, C. (1985). Spectral integration based on common amplitude modulation. *Perception & Psychophysics*, 37, 483–493.
- Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, 89(2), 244–249.
- Bregman, A. S., Liao, C., & Levitan, R. (1990). Auditory grouping based on fundamental frequency and formant peak frequency. *Canadian Journal of Psychology*, 44, 400–413.
- Bregman, A. S., & Pinker, S. (1978). Auditory streaming and the building of timbre. *Canadian Journal of Psychology*, 32(1), 19–31.
- Bregman, A. S., & Doehring, P. (1984). Fusion of simultaneous tonal glides: The role of parallelness and simple frequency relations. *Perception & Psychophysics*, 36(3), 251–256.
- Carlsen, J. C. (1981). Some factors which influence melodic expectancy. *Psychomusicology*, 1, 12–29.
- Carlyon, R. P. (1991). Discriminating between coherent and incoherent frequency modulation of complex tones. *Journal of the Acoustical Society of America*, 89(1), 329–340.
- Carlyon, R. P. (1992). The psychophysics of concurrent sound segregation. In R. P. Carlyon, C. J. Darwin, & I. J. Russell (Eds.), *Processing of Complex Sounds by the Auditory System* (pp. 53–61). Oxford: Clarendon Press/Oxford University Press.
- Carlyon, R. P. (1994). Further evidence against an across-frequency mechanism specific to the detection of frequency modulation (FM) incoherence between resolved frequency components. *Journal of the Acoustical Society of America*, 95(2), 949–961.
- Cazden, N. (1945). Musical consonance and dissonance: A cultural criterion. *Journal of Aesthetics and Art Criticism*, 4(1), 3–11.
- Chowning, J. M. (1980). Computer synthesis of the singing voice. In: J. Sundberg (Ed.) *Sound generation in winds, strings, computers* (pp. 4–13). Stockholm: Royal Swedish Academy of Music.
- Ciocca, V., & Bregman, A. S. (1987). Perceived continuity of gliding and steady-state tones through interrupting noise. *Perception & Psychophysics*, 42, 476–484.
- Crowder, R. G. (1969). Improved recall for digits with delayed recall cues. *Journal of Experimental Psychology* 82, 258–262.
- Darwin, C. J., & Gardner, R. B. (1986). Mistuning a harmonic of a vowel: Grouping and phase effects on vowel quality. *Journal of the Acoustical Society of America*, 79, 838–845.
- Descocudres, A. (1921). *Le développement de l'enfant de deux à sept ans*. Paris: Delachaux & Niestle.
- Deutsch, D. (1975). Two-channel listening to musical scales. *Journal of the Acoustical Society of America*, 57, 1156–1160.
- DeWitt, L. A., & Crowder, R. G. (1987). Tonal fusion of consonant musical intervals: The oomph in Stumpf. *Perception & Psychophysics*, 41(1), 73–84.
- Divenyi, P. L. (1993). Auditory organization in music: Let the (defunct) composer speak. *Journal of the Acoustical Society of America*, 93(4), S2362.
- Divenyi, P. L., & Oliver, S. K. (1989). Resolution of steady-state sounds in simulated auditory space. *Journal of the Acoustical Society of America*, 85, 2042–2052.

- Dowling, W. J. (1967). *Rhythmic fission and the perceptual organization of tone sequences*. Unpublished doctoral dissertation, Harvard University, Cambridge, MA.
- Dowling, W. J. (1973). The perception of interleaved melodies. *Cognitive Psychology*, 5, 322–337.
- Duda, R. O., & Hart, P. E. (1973). *Pattern classification and scene analysis*. New York: John Wiley & Sons.
- Erickson, R. (1975). *Sound structure in music*. Berkeley: University of California Press.
- Estes, B., & Combes, A. (1966). Perception of quantity. *Journal of Genetic Psychology*, 108, 333–336.
- Fétis, F.-J. (1840). *Esquisse de l'histoire de l'harmonie*. Paris: Revue et Gazette Musicale de Paris. [English trans. by M. Arlin, Stuyvesant, NY: Pendragon Press, 1994.]
- Fitts, P. M. (1954). The information capacity of the human motor system in controlling amplitude of movement. *Journal of Experimental Psychology*, 47, 381–391.
- Fitzgibbons, P. J., Pollatsek, A., & Thomas, I. B. (1974). Detection of temporal gaps within and between perceptual tonal groups. *Perception & Psychophysics*, 16(3), 522–528.
- Fletcher, H. (1940). Auditory patterns. *Reviews of Modern Physics*, 12, 47–65.
- Fletcher, H. (1953). *Speech and hearing in communication*. New York: Van Nostrand.
- Fux, J. J. (1725). *Gradus ad Parnassum*. Vienna. [Trans. by A. Mann as *The study of counterpoint*. New York: Norton Co., 1971.]
- Gauldin, R. (1985). *A practical approach to sixteenth-century counterpoint*. Englewood Cliffs, NJ: Prentice-Hall.
- Gelman, R., & Tucker, M. F. (1975). Further investigations of the young child's conception of number. *Child Development*, 46, 167–175.
- Gjerdingen, R. O. (1994). Apparent motion in music? *Music Perception*, 11(4), 335–370.
- Glasberg, B. R., & Moore, B. C. J. (1990). Derivation of auditory filter shapes from notched noise data. *Hearing Research*, 47, 103–138.
- Glucksberg, S., & Cowen, G. N., Jr. (1970). Memory for nonattended auditory material. *Cognitive Psychology*, 1, 149–156.
- Greenwood, D. D. (1961a). Auditory masking and the critical band. *Journal of the Acoustical Society of America*, 33(4), 484–502.
- Greenwood, D. D. (1961b). Critical bandwidth and the frequency coordinates of the basilar membrane. *Journal of the Acoustical Society of America*, 33(4), 1344–1356.
- Greenwood, D. D. (1990). A cochlear frequency-position function for several species—29 years later. *Journal of the Acoustical Society of America*, 87(6), 2592–2605.
- Greenwood, D. D. (1991). Critical bandwidth and consonance in relation to cochlear frequency-position coordinates. *Hearing Research*, 54(2), 164–208.
- Gregory, A. H. (1994). Timbre and auditory streaming. *Music Perception*, 12, 161–174.
- Guttman, N., & Julesz, B. (1963). Lower limits of auditory periodicity analysis. *Journal of the Acoustical Society of America*, 35, 610.
- Haas, H. (1951). Über den Einfluss eines Einfachechos auf die Hörsamkeit von Sprache. *Acustica*, 1, 49–58.
- Hartmann, W. M., & Johnson, D. (1991). Stream segregation and peripheral channeling. *Music Perception*, 9(2), 155–183.
- Heise, G. A., & Miller, G. A. (1951). An experimental study of auditory patterns. *American Journal of Psychology*, 64, 68–77.
- Helmholtz, H. L. von (1863/1878). *Die Lehre von der Tonempfindungen als physiologische Grundlage für die Theorie der Musik*. Braunschweig: Verlag F. Vieweg & Sohn. [Translated as: *On the sensations of tone as a physiological basis for the theory of music* (1878); second English edition, New York: Dover Publications, 1954.]
- Hindemith, P. (1944). *A concentrated course in traditional harmony* (2 vols., 2nd ed.). New York: Associated Music Publishers.
- von Hippel, P. (2000). Redefining pitch proximity: Tessitura and mobility as constraints on melodic interval size. *Music Perception*, 17(3), 315–327.
- von Hippel, P. & Huron, D. (2000). Why do skips precede reversals? The effect of tessitura on melodic structure. *Music Perception*, 18(1), 59–85.

- Hirsh, I. J. (1959). Auditory perception of temporal order. *Journal of the Acoustical Society of America*, 31(6), 759–767.
- Horwood, F. J. (1948). *The basis of harmony*. Toronto: Gordon V. Thompson.
- Houtgast, T. (1971). Psychophysical evidence for lateral inhibition in hearing. *Proceedings of the 7th International Congress on Acoustics (Budapest)*, 3, 24 H15, 521–524.
- Houtgast, T. (1972). Psychophysical evidence for lateral inhibition in hearing. *Journal of the Acoustical Society of America*, 51, 1885–1894.
- Houtgast, T. (1973). Psychophysical experiments on “tuning curves” and “two-tone inhibition.” *Acustica*, 29, 168–179.
- Huron, D. (1987). *Auditory stream segregation and voice independence in multi-voice musical textures*. Paper presented at the Second Conference on Science and Music, City University, London, UK.
- Huron, D. (1989a). *Voice segregation in selected polyphonic keyboard works by Johann Sebastian Bach*. Unpublished doctoral dissertation, University of Nottingham, Nottingham, England.
- Huron, D. (1989b). Voice denumerability in polyphonic music of homogeneous timbres. *Music Perception*, 6(4), 361–382.
- Huron, D. (1989c). Characterizing musical textures. *Proceedings of the 1989 International Computer Music Conference* (pp. 131–134). San Francisco: Computer Music Association.
- Huron, D. (1991a). The avoidance of part-crossing in polyphonic music: perceptual evidence and musical practice. *Music Perception*, 9(1), 93–104.
- Huron, D. (1991b). Tonal consonance versus tonal fusion in polyphonic sonorities. *Music Perception*, 9(2), 135–154.
- Huron, D. (1991c). Review of the book *Auditory scene analysis: The perceptual organization of sound*. *Psychology of Music*, 19(1), 77–82.
- Huron, D. (1993a). Note-onset asynchrony in J. S. Bach’s two-part Inventions. *Music Perception*, 10(4), 435–443.
- Huron, D. (1993b). Chordal-tone doubling and the enhancement of key perception. *Psychomusicology*, 12(1), 73–83.
- Huron, D. (1993c). A derivation of the rules of voice-leading from perceptual principles. *Journal of the Acoustical Society of America*, 93(4), S2362.
- Huron, D. (1994). Interval-class content in equally-tempered pitch-class sets: Common scales exhibit optimum tonal consonance. *Music Perception*, 11(3), 289–305.
- Huron, D. (1997). *Stream segregation in polyphonic music: Asynchronous preparation of tonally fused intervals in J. S. Bach*. Unpublished manuscript.
- Huron, D., & Fantini, D. (1989). The avoidance of inner-voice entries: Perceptual evidence and musical practice. *Music Perception*, 7(1), 43–47.
- Huron, D., & Mondor, T. (1994). *Melodic line, melodic motion, and Fitts’ law*. Unpublished manuscript.
- Huron, D., & Parncutt, R. (1992). *How “middle” is middle C? Terhardt’s virtual pitch weight and the distribution of pitches in music*. Unpublished manuscript.
- Huron, D., & Parncutt, R. (1993). An improved model of tonality perception incorporating pitch salience and echoic memory. *Psychomusicology*, 12(2), 154–171.
- Huron, D., & Sellmer, P. (1992). Critical bands and the spelling of vertical sonorities. *Music Perception*, 10(2), 129–149.
- Iverson, P. (1992). Auditory stream segregation by musical timbre. Paper presented at the 123rd meeting of the Acoustical Society of America. *Journal of the Acoustical Society of America*, 91(2), Pt. 2, 2333.
- Iyer, I., Aarden, B., Hoglund, E., & Huron, D. (1999). Effect of intensity on sensory dissonance. *Journal of the Acoustical Society of America*, 106(4), 2208–2209.
- Jeppesen, K. (1939/1963). *Counterpoint: The polyphonic vocal style of the sixteenth century* (G. Haydon. Transl.). Englewood Cliffs, NJ: Prentice-Hall, 1963. (Original work published 1931)
- Kaestner, G. (1909). Untersuchungen über den Gefühlseindruck unanalysierter Zweiklänge. *Psychologische Studien*, 4, 473–504.

- Kameoka, A., & Kuriyagawa, M. (1969a). Consonance theory. Part I: Consonance of dyads. *Journal of the Acoustical Society of America*, 45, 1451–1459.
- Kameoka, A., & Kuriyagawa, M. (1969b). Consonance theory. Part II: Consonance of complex tones and its calculation method. *Journal of the Acoustical Society of America*, 45, 1460–1469.
- Keys, I. (1961). *The texture of music: From Purcell to Brahms*. London: Dennis Dobson.
- Körte, A. (1915). Kinomatoscopische Untersuchungen. *Zeitschrift für Psychologie der Sinnesorgane*, 72, 193–296.
- Kringlebotn, M., Gundersen, T., Krokstad, A., & Skarstein, O. (1979). Noise-induced hearing losses. Can they be explained by basilar membrane movement? *Acta Oto-Laryngologica Supplement*, 360, 98–101.
- Krumhansl, C. (1979). The psychological representation of musical pitch in a tonal context. *Cognitive Psychology*, 11, 346–374.
- Kubovy, M., & Howard, F. P. (1976). Persistence of pitch-segregating echoic memory. *Journal of Experimental Psychology: Human Perception & Performance*, 2(4), 531–537.
- Lachenbruch, P. A. (1975). *Discriminant Analysis*. New York: Hafner Press.
- Martin, D. (1965). The evolution of barbershop harmony. *Music Journal Annual*, p. 40.
- Mayer, M. (1894). Researches in Acoustics. No. IX. *Philosophical Magazine*, 37, 259–288.
- McAdams, S. (1982). Spectral fusion and the creation of auditory images. In M. Clynes (Ed.), *Music, mind, and brain: The neuropsychology of music* (pp. 279–298). New York & London: Plenum Press.
- McAdams, S. (1984a). Segregation of concurrent sounds. I: Effects of frequency modulation coherence. *Journal of the Acoustical Society of America*, 86, 2148–2159.
- McAdams, S. (1984b). *Spectral fusion, spectral parsing and the formation of auditory images*. Unpublished doctoral dissertation, Stanford University, Palo Alto, CA.
- McAdams, S., & Bregman, A. S. (1979). Hearing musical streams. *Computer Music Journal*, 3(4), 26–43, 60, 63.
- Merriam, A. P., Whinery, S., & Fred, B. G. (1956). Songs of a Rada community in Trinidad. *Anthropos*, 51, 157–174.
- Miller, G. A., & Heise, G. A. (1950). The trill threshold. *Journal of the Acoustical Society of America*, 22(5), 637–638.
- Moore, B. C. J., & Glasberg, B. R. (1983). Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *Journal of the Acoustical Society of America*, 74(3), 750–753.
- Morris, R. O. (1946). *The Oxford harmony*. Vol. 1. London: Oxford University Press.
- Mursell, J. (1937). *The psychology of music*. New York: Norton.
- Narmour, E. (1991). *The Analysis and cognition of basic melodic structures: The implication-realization model*. Chicago: University of Chicago Press.
- Narmour, E. (1992). *The analysis and cognition of melodic complexity: The implication-realization model*. Chicago: University of Chicago Press.
- Neisser, U. (1967). *Cognitive psychology*. New York: Meredith.
- van Noorden, L. P. A. S. (1971a). Rhythmic fission as a function of tone rate. *IPO Annual Progress Report*, 6, 9–12.
- van Noorden, L. P. A. S. (1971b). Discrimination of time intervals bounded by tones of different frequencies. *IPO Annual Progress Report*, 6, 12–15.
- van Noorden, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences*. Doctoral dissertation, Technisch Hogeschool Eindhoven; published Eindhoven: Druk van Voorschoten.
- Norman, D. (1967). Temporal confusions and limited capacity processors. *Acta Psychologica*, 27, 293–297.
- Ohgushi, K., & Hato, T. (1989). On the perception of the musical pitch of high frequency tones. *Proceedings of the 13th International Congress on Acoustics (Belgrade)*, 3, 27–30.
- Opolko, F., & Wapnick, J. (1989). *McGill University Master Samples*. Compact discs Vols. 1–11. Montreal: McGill University, Faculty of Music.

- Ortmann, O. R. (1926). *On the melodic relativity of tones*. Princeton, NJ: Psychological Review Company. (Vol. 35, No. 1 of Psychological Monographs.)
- Parks, R. (1984). *Eighteenth-century counterpoint and tonal structure*. Englewood Cliffs, NJ: Prentice-Hall.
- Parncutt, R. (1989). *Harmony: A psychoacoustical approach*. Berlin: Springer Verlag.
- Parncutt, R. (1993). Pitch properties of chords of octave-spaced tones. *Contemporary Music Review*, 9, 35–50.
- Pattison, H., Gardner, R. B., & Darwin, C. J. (1986). Effects of acoustical context on perceived vowel quality. *Journal of the Acoustical Society of America*, 80, Suppl. 1.
- Piston, W. (1978). *Harmony*. 4th ed. New York: W.W. Norton & Co.
- Plomp, R., & Levelt, W. J. M. (1965). Tonal consonance and critical bandwidth. *Journal of the Acoustical Society of America*, 37, 548–560.
- Plomp, R., & Steeneken, H. J. M. (1968). Interference between two simple tones. *Journal of the Acoustical Society of America*, 43, 883–884.
- Price, P. (1984). Carillon. In S. Sadie (Ed.), *New Grove dictionary of musical instruments* (Vol. 1, pp. 307–311). London: Macmillan Press.
- Rameau, J.-P. (1722). *Traité de l'harmonie*. Paris: Jean Baptiste Christophe Ballard. [Facsimile reprint, New York: Broude Bros., 1965.]
- Rasch, R. A. (1978). The perception of simultaneous notes such as in polyphonic music. *Acustica*, 40, 21–33.
- Rasch, R. A. (1979). Synchronization in performed ensemble music. *Acustica*, 43, 121–131.
- Rasch, R. A. (1981). *Aspects of the perception and performance of polyphonic music*. Doctoral dissertation. Utrecht, The Netherlands: Elinkwijk BV.
- Rasch, R. A. (1988). Timing and synchronization in ensemble performance. In J. Sloboda (Ed.), *Generative processes in music: The psychology of performance, improvisation, and composition* (pp. 70–90). Oxford: Clarendon Press.
- Riemann, H. (1903). *Grosse Kompositionslehre*. Berlin and Stuttgart: W. Spemann.
- Root, D. L. (1980). Barbershop harmony. In S. Sadie (Ed.), *New Grove Dictionary of Music and Musicians* (Vol. 2, p. 137). London: Macmillan.
- Rössler, E. K. (1952). *Klangfunktion und Registrierung*. Kassel: Bärenreiter Verlag.
- Rostron, A. B. (1974). Brief auditory storage: Some further observations. *Acta Psychologica*, 38, 471–482.
- Saldanha, E. L., & Corso, J. F. (1964). Timbre cues and the identification of musical instruments. *Journal of the Acoustical Society of America*, 36, 2021–2026.
- Sandell, G. J. (1991a). A library of orchestral instrument spectra. *Proceedings of the 1991 International Computer Music Conference*, 98–101.
- Sandell, G. J. (1991b). *Concurrent timbres in orchestration: A perceptual study of factors determining "blend."* Unpublished doctoral dissertation, Northwestern University, Evanston, Illinois.
- Schaeffer, B., Eggleston, V. H., & Scott, J. L. (1974). Number development in young children. *Cognitive Psychology*, 6, 357–379.
- Scharf, B. (1961). Loudness summation under masking. *Journal of the Acoustical Society of America*, 33(4), 503–511.
- Schenker, H. (1906). *Neue musikalische Theorien und Phantasien*. Vol. 1: *Harmonielehre*. Stuttgart & Berlin: J.G. Cotta'sche Buchhandlung Nachfolger.
- Schoenberg, A. (1991/1978). *Harmonielehre*. Vienna: Universal Edition. [R. E. Carter, Transl., *Theory of Harmony*. Berkeley: University of California Press, 1978.]
- Schouten, J. F. (1940). The residue, a new component in subjective sound analysis. *Proceedings Koninklijke Nederlandse Akademie Wetenschappen*, 43, 356–365.
- Schouten, J. F. (1962). On the perception of sound and speech. *Proceedings of the 4th International Congress on Acoustics (Copenhagen)*, 2, 201–203.
- Semal, C., & Demany, L. (1990). The upper limit of "musical" pitch. *Music Perception*, 8(2), 165–175.
- Shepard, R. N. (1981). Psychophysical complementarity. In M. Kubovy, J. Pomerantz (Eds.), *Perceptual organization*. Hillsdale, NJ: Lawrence Erlbaum Associates, 279–341.

- Stainer, J. (1878). *Harmony*. London: Novello and Co.
- Steiger, H., & Bregman, A. S. (1981). Capturing frequency components of glided tones: Frequency separation, orientation and alignment. *Perception & Psychophysics*, 30, 425-435.
- Strauss, M. S., & Curtis, L. E. (1981). Infant perception of numerosity. *Child Development*, 52(4), 1146-1152.
- Strong, W., & Clark, M. (1967). Synthesis of wind-instrument tones. *Journal of the Acoustical Society of America*, 41(1), 39-52.
- Stumpf, C. (1890). *Tonpsychologie*. Vol. 2. Leipzig: S. Hirzel.
- Sundberg, J., Askenfelt, A., & Frydén, L. (1983). Music performance: A synthesis-by-rule approach. *Computer Music Journal*, 7(1), 37-43.
- Terhardt, E., Stoll, G., & Seewann, M. (1982a). Pitch of complex signals according to virtual-pitch theory: Tests, examples, and predictions. *Journal of the Acoustical Society of America*, 71(3): 671-678.
- Terhardt, E., Stoll, G., & Seewann, M. (1982b). Algorithm for extraction of pitch and pitch salience from complex tonal signals. *Journal of the Acoustical Society of America*, 71(3): 679-688.
- Terhardt, E., Stoll, G., Schermbach, R., & Parncutt, R. (1986). Tonhöhenmehrdeutigkeit, Tonverwandtschaft und Identifikation von Sukzessivintervallen. *Acustica*, 61, 57-66.
- Thurlow, W. R. (1957). An auditory figure-ground-effect. *American Journal of Psychology*, 70, 653-654.
- Tougas, Y., & Bregman, A. S. (1985). Crossing of auditory streams. *Journal of Experimental Psychology: Human Perception and Performance*, 11(6), 788-798.
- Treisman, A. M. (1964). Monitoring and storage of irrelevant messages in selective attention. *Journal of Verbal Learning and Verbal Behavior*, 3, 449-459.
- Treisman, A. M., & Howarth, C. I. (1959). Changes in threshold level produced by a signal preceding or following the threshold stimulus. *Quarterly Journal of Experimental Psychology*, 11, 129-142.
- Treisman, A. M., & Rostron, A. B. (1972). Brief auditory storage: A modification of Sperling's paradigm applied to audition. *Acta Psychologica*, 36, 161-170.
- Trythall, G. (1993). *Eighteenth century counterpoint*. Madison, WI: Brown and Benchmark.
- Vicario, G. (1960). Analisi sperimentale di un caso di dipendenza fenomenica tra eventi sonori. *Rivista di psicologia*, 54(3), 83-106.
- Vos, J. (1995). Perceptual separation of simultaneous complex tones: the effect of slightly asynchronous onsets. *Acta Acustica*, 3, 405-416.
- Warren, R. M., Obusek, C. J., & Ackroff, J. (1972). Auditory induction: Perceptual synthesis of absent sounds. *Science*, 176, 1149-1151.
- Wessel, D. L. (1979). Timbre space as a musical control structure. *Computer Music Journal*, 3(2), 45-52.
- Wright, J. K. (1986). Auditory object perception: Counterpoint in a new context. Unpublished masters thesis, McGill University, Montréal, Québec.
- Wright, J. K., & Bregman, A. S. (1987). Auditory stream segregation and the control of dissonance in polyphonic music. *Contemporary Music Review*, 2, 63-93.
- Zwicker, E.G., Flottorp, G., & Stevens, S. S. (1957). Critical bandwidth in loudness summation. *Journal of the Acoustical Society of America*, 29, 548-557.