Review: David Huron, *Voice Leading: The Science Behind a Musical Art* (Cambridge, MIT Press, 2016) viii + 263 pp. ISBN: 9780262034852

Gilberto Bernardes

INESC TEC

gba@inesctec.pt

THE ART OF COMBINING CONCURRENT MUSICAL LINES, or voice leading, has been an instrumental concept in Western music and the subject of considerable discussion over the last centuries. Traditionally, voice leading is understood, discussed, and taught as a set of dogmatic canons (i.e., *dos* and *don'ts*) based on generalizations inferred from musical practice. Conversely, David Huron, at Ohio State University, has pursued during his remarkable academic career a musicological study on voice leading from perceptual and cognitive standpoints, which he has now synthesized in a lucid book.[1]

Huron's skeptical attitude towards the dogmatic nature of the longstanding part-writing rule-based tradition has led to some of the most insightful considerations over recent decades on voice leading, and sheds light on the explicit relation between perception, cognition, and part-writing rules. In other words, Huron provides a scientific explanation that demonstrates in timely fashion why particular music textures are more compelling to the human ear. However, he positions the practice as being driven by a multitude of factors, notably including cultural, idiomatic, and biological (e.g., perceptual and cognitive) traits.

---

[1] The book under review is a mature and expanded version of a previous publication (HURON 2001), which already draws a solid and compound framework for the perceptual and cognitive basis underlying canonic part-writing rules.

Under the umbrella of two main scientific domains, hearing sciences—particularly auditory scene analysis[2]—and empirical musicology,[3] Huron draws from a music theorist's perspective a compelling argumentation concerning the fundamental principles of the auditory human system which govern the perception of musical textures. With an accessible, yet scientifically rigorous tone, he unpacks the mostly bottom-up mechanics by which humans create an 'auditory image' in the brain from an 'acoustic' (or physical) 'source'. Particular attention is given to the processes through which our brains 'integrate' a complex harmonic tone into a unique image as well as the way in which 'segregated' auditory images are resolved from a complex auditory source.

After an introductory chapter that compiles a set of sixteen canonic voice-leading rules from the late Baroque period,[4] Huron delves into to the auditory principles that are relevant from the standpoint of a cognitive theory of musical texture. These principles are addressed in the book in order to support an argumentation which ultimately explains how two or more concurrent parts, voices, or textures can be clearly perceived as independent streams, as well as how they serve particular musical goals, if those goals are worthwhile, and even effective ways of achieving them.

In total, twelve principles from auditory scene analysis are evoked: toneness, harmonic fusion, auditory masking, continuity, pitch proximity, pitch co-modulation, onset asynchrony, limited density, timbral differentiation, source location, attention, and expectation. The order of the principles follows their presentation in the book, which itself relates to their generalized adoption in musical part-writing. More specifically, while the first six principles are prominent in part-writing, the remaining six do not show a consistent adoption. The principle and their influence over the musical structure are presented afterwards.

Empirical research has shown that high degrees of toneness[5] evoke better auditory images, supporting the dominance of instruments producing complex harmonic tones in part-writing over those

---

[2] 'Auditory scene analysis' is a term coined by the psychologist Albert BREGMAN (1990) to address the mechanics of the human auditory system in parsing complex auditory environments into meaningful elements. Its underlying aim is to explain concurrently how we 'segregate' different sound sources and 'integrate' the components of a sound into coherent auditory images from complex 'auditory scenes'. The percept of a group of sequential and/or simultaneous sounds as a coherent whole appearing to come from a single sound source is known as 'auditory stream' or, simply, 'stream'.

[3] Musicological studies are always to some extent 'empirical', in that they assess how a given theory, idea or interpretation conforms to particular musical works. However, in the current context, empirical musicology denotes the branch of musicology that follows empirical approaches typically applied in social sciences, towards an understanding of the music as a complex practice which results from multiple viewpoints, particularly including music theory, computational musicology, ethnomusicology, psychology and sociology of music, data analysis, and statistics.

[4] A rationale for the preference of part-writing rules of the Baroque period is fully detailed in chapter 14 (pp. 179-82).

[5] Toneness is a psychoacoustic metric that aims to capture the degree of clarity of pitch perceptions (PARNCUTT 1989).

which produce inharmonic (or aperiodic) sounds, such as wood blocks, tambourine, and cymbals. Additionally, the E2 - G5[6] (82.4 - 783.9 Hz) pitch range is known to produce clearer images and less ambiguous pitches (TERHARDT - STOLL - SEEWANN 1982a, 1982b). This thus explains, in Huron's view, why most musical instruments fit into this range.

While attempting to conform all partials stimulating the basilar membrane into (hypothetical) harmonic templates, our brain can fuse[7] distinct tones into a single auditory image. The better a set of partials conforms to a harmonic series, the higher the likelihood of it being fused into a single image. In the case of two complex harmonic tones, if the fundamental frequencies are related by simple integer ratios, this phenomenon tends to be more prominent.[8]

Auditory masking is a phenomenon that occurs in the brain by which frequencies with higher amplitude within a critical bandwidth obscure others. Empirical evidence shows that chord spacings are consistent with minimal masking,[9] or, in other words, with an even distribution of activation points along the basilar membrane. This interaction is also known to vary with the register, equating to wider intervals in the bass than in the treble regions.[10] Furthermore, the predominance of the highest voice in a musical texture can also be explained by this phenomenon, as higher harmonic tones tend to mask more predominantly the partials of lower tones (CHON - HURON 2014). Highly connected with the concept of auditory masking is sensory irritation, which is created by the inability of the auditory system fully to resolve two frequencies within a critical bandwidth. This inability is shown to be, together with other factors, involved in the perception of dissonance.

---

[6] I have adopted the octave index 4 as the middle octave of the scale and the standard tuning pitch is A4 (A above middle C) 440 Hertz(Hz).

[7] Tonal fusion is the tendency for some concurrent sounds 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). The likelihood of two tones fusing increases if their fundamental frequencies are related by simple integer ratios and their spectral content conform to a single (hypothetical) harmonic series.

[8] The most fused intervals in a decreasing order are unison, octave, fifteenth (double octave), perfect twelfth (an octave plus a perfect fifth), and perfect fifth. Note that the suggested ranking order of the most fused tones is prosed by Huron in the book under review, yet there is no general agreement in the literature concerning this ranking. HURON (1991) includes a comprehensive discussion on this topic.

[9] HURON (2001) showed that composers consistently space the component notes of chords evenly across the basilar membrane. This suggests that composers tend to minimize auditory masking which itself avoids sensory irritation. This explains, for example, why the spacing between bass and tenor voices tends to be wider than the other successive voice combinations.

[10] Above middle C, the region of interference ('critical band') is roughly a minor third. Below middle C this region of interference increases in inverse proportion to lower pitch roughly reaching an augmented fifth at the bottom of the bass staff. Above the middle C the regions become progressively smaller, approaching a major second at the top of the treble staff.

After clarifying the principles of static image formation, Huron introduces three other principles which are particularly relevant to understand how different streams are created in time—continuity, pitch proximity, and pitch co-modulation principles. The continuity and proximity (intervallic distance) between tones is shown to affect how we connect two tones, and thus ensures the integration of dynamic sources into a single stream. The smaller the gap (or silence) and intervallic distance between them, the better the connection is established or the stream formed. Two boundaries have been found to explain stream formation. When the tempo is slow and/or the pitches distances are small, a single stream is perceived (below the so-called 'trill boundary'). Conversely, when the tempo is fast and the pitch distances are large, two streams are always perceived (V̲AN N̲OORDEN 1975). Furthermore, perceptual union is enhanced when tones move in parallel motion or co-modulate (i.e., in a similar pitch direction). In musical practice, pitch co-modulation can serve different purposes. If a composer wants to enhance the segregation of concurrent parts, parallel motion should be avoided. If a composer wants to integrate different parts parallel motion is recommended, as found in the works of Pyotr Tchaikovsky ('Dance of the Reed Pipes' from the *Nutcracker*) or Claude Debussy (*La Cathédrale engloutie*).

The onset asynchrony principle explains how stream segregation can be enhanced depending on the temporal gap separating the onsets of each musical part.[11] If a composer aims at music in which the parts have a high degree of perceptual independence (particularly important when in the presence of intervals susceptible of harmonic fusion, such as octaves and fifths), then onset asynchrony is recommended (or synchronous note onsets ought to be avoided). While a vast array of musical works conforms to this principle, many exceptions can be identified. Of note are the choral works of Johann Sebastian Bach, which are prototypical works in obeying most perceptual principles, while simultaneously contradicting the onset asynchrony principle. As a means to enhance compositional dimensions such as the intelligibility of the lyrics or the creation of rhythmic momentums, Huron argues that composers can compromise the onset asynchrony principle.

Perceptual experiments have shown that humans have limited capacity to track concurrent streams if they exceed three layers,[12] even when performed by different instruments (R̲ASCH 1981), thus reinforcing the natural intuition of musicians to restrict the number of voices of polyphonic music in agreement with the perceptual experiments. However, exceptions to this rule are easy to identify. For

---

[11] On average, a minimum of 80 milliseconds is required for two percussive sounds not to fuse into a single auditory image. This time increases to 100 milliseconds for less percussive sounds.

[12] Huron draws a careful distinction between a musical notated part and an auditory stream, in that a notated part does not necessarily evoke an auditory stream.

example, contrary to the three streams limit principle, most music theory books on tonal harmony tend to focus on four-part writing. Huron justifies the popularity of such choice as a compositional compromise between perceptual principles and the extended possibilities offered by four-part writing in adopting a wider range of chord types.

Empirical experiments, in agreement with a large corpus of music, have shown that timbre is important in stream formation, and even takes precedence over pitch-based streaming (MCADAMS - BREGMAN 1979) for integrating dynamic streams and segregating concurrent streams. Yet, once again Huron points out how this principle is often ignored in Western music, for example string quartets, choirs, and piano. This has happened for a number of historical and practical reasons, including the difficulties in recruiting musicians, balancing the disparities of instrumental timbre and volume, or simply because of the preference for blended sonorities.

The displacement of sound sources is known to be one of the most effective ways of providing enhanced stream segregation (DIVENYI - OLIVER 1989). While this principle is quite evident in electroacoustic works, it is followed to a much smaller degree in previous periods (with the remarkable exception of the antiphonal works of Giovanni Gabrieli), which Huron relates to practical reasons whereby musicians need good visual and auditory feedback to perform.

The concept of 'attention' is a perceptual principle introduced by Huron to explain how, among many other perceptual principles, embellishing tones can enhance a particular auditory stream by introducing a solo event. This is shown to be consistent among a large corpus of Bach's polyphonic works (HURON 2007). Later in the book, Huron revisits the principle of attention to explain how listeners can shift their focus to different levels of the hierarchical musical structure (HURON 2007, 13).

A sole exception to the bottom-up principles in auditory image formation acknowledged by Huron is the concept of expectation, which accounts for the feeling of direction (leading somewhere) of the musical discourse formed by simple exposure (i.e., 'implicit' or 'statistical learning'). A reasoning based on two forms of musical expectations is described to account for the sense of poor voice leading which results from reversing a musical excerpt, while, nonetheless, conforming to most of the remaining part-writing perceptual principles.[13] Specifically, when reversing a piece of music both 'schematic' (conventional musical patterns, such as prototypical harmonic sequences, e.g. V-I) and

---

[13] The concept of expectation is presented and developed in the book under review in order to illustrate how the psychological mechanisms of expectation work in the case of part-writing. The ground theory supporting the connection was previously presented in Huron's book *Sweet Anticipation* (HURON 2006), in which he set up a general theory of music perception and cognition, with expectation as the central concept.

veridical expectations (quotations or allusions to 'familiar' melodies or specific pieces) are violated. Huron additionally evokes a third form of 'dynamic-adaptive' expectations, which relate to the anticipation of contextual patterns within a composition, such as repetitive or melodic patterns (e.g. recurring figures, themes and motives typically featured in fugues). Therefore, expectations not only help us in tracking different streams, but also contribute to the sense of directionality, thus transforming 'good part writing into good voice-leading' (p. 147).

In the final chapters of the book (pp. 9-13), Huron inverts the reasoning used so far, interpreting canonic part-writing rules as the interplay of several perceptual principles, including chordal-tone doublings (p. 11), direct intervals (p. 12), musical hierarchies above tone objects (p. 13), and even to reinforce the understanding of embellishing tones (p. 9). Yet, while a careful treatment of how compositional goals may circumvent particular principles, deeper considerations on the explicit interaction between the stated perceptual principles—which to some extent can 'cancel' each other[14]— is surprisingly absent.

The concluding chapter (p. 17) reviews, 'descriptively' rather than 'prescriptively', a reformulated set of voice leading preference rules driven from perceptual principles.[15] In identifying a conceptual framework for understanding musical textures, without providing a novel methodological plan, Huron devises new strategies for the analyst and the composer. A fertile terrain for understanding music from an overlooked perspective is set, which seems particularly relevant to the provision of alternative approaches to the analyis of some twentieth-century repertoire for which formalized methods are still lacking. Repertoire whose primarily building blocks of musical structure are not pitch and duration benefit particularly from a perceptually-grounded music analysis. Typical examples of such works are the sonic masses in Iannis Xenakis (*Persephassa* and *La Légende d'Eer*), György Ligeti (*Atmosphères* and *Lontano*), and Tristan Murail (*Gondwana* and *Désintégrations*).[16]

In light of the perceptual principles identified, Huron sets out the groundwork for the creation of formal analytical and compositional methods, notably for electroacoustic works that target microsonic

---

[14] Besides the interaction between harmonic fusion and onset synchrony (pp. 106-7), no other explicit interaction of perceptual principles is discussed at any length.

[15] Despite the compelling argumentation presented in the book, Huron shows evidence that these perceptual principles are neither universal nor innate (pp. 185-93), and result from, or at least are susceptible to, experience. Moreover, he draws attention to the temptation of using them as compositional imperatives.

[16] The analysis of music outside of the Western baroque canon can prove quite successful in establishing how the perceptual principles evoked interact. A particularly simple example is the examination of the behaviour of the principles in 'spectral' music in light of the expected maximal harmonic fusion.

structures. This outstanding book enables the creation of tools for analysing and designing musical scenes beyond the discretized canonic part-writing rules and the confined universe of canonical Western instrumental music.

References

BREGMAN, Albert S. (1990), *Auditory Scene Analysis: The Perceptual Organization of Sound* (Cambridge - MA, MIT press)

CHON, Song H. and David HURON (2014), 'Does Auditory Masking Explain High Voice Superiority', *Proceedings of the 13th International Conference of Music Perception and Cognition (ICMPC)*, pp. 185-8

DEWITT, Lucinda A. and Robert G. CROWDER (1987), 'Tonal Fusion of Consonant Musical Intervals: The Oomph in Stumpf', *Perception & Psychophysics*, 41/1, pp. 73-84

DIVENYI, Pierre L. and Susan K. OLIVER (1989), 'Resolution of Steady-state Sounds in Simulated Auditory Space', *The Journal of the Acoustical Society of America*, 85/5, pp. 2042-52

HURON, David (1991), 'Tonal Consonance Versus Tonal Fusion in Polyphonic Sonorities', *Music Perception*, 9/2, pp. 135-54

HURON, David (2001), 'Tone and Voice: A Derivation of the Rules of Voice-leading from Perceptual Principles', *Music Perception*, 19/1, pp. 1-64

HURON, David (2006), *Sweet Anticipation: Music and the Psychology of Expectation* (Cambridge - MA, MIT press)

HURON, David (2007), 'On the Role of Embellishment Tones in the Perceptual Segregation of Concurrent Musical Parts', *Empirical Musicology Review*, 2/4, pp. 123-39

MCADAMS, Stephen and Albert BREGMAN (1979), 'Hearing Musical Streams', *Computer Music Journal*, pp. 26-60

PARNCUTT, Richard (1989), *Harmony: A Psychoacoustic Approach* (Berlin, Springler-Verlag)

RASCH, Rudolf A. (1981), *Aspects of the Perception and Performance of Polyphonic Music* (PhD dissertation, Utrecht, Elinkwijk BV)

TERHARDT, Ernst, Gerhard STOLL and Manfred SEEWANN (1982a), 'Algorithm for Extraction of Pitch and Pitch Salience from Complex Tonal Signals', *The Journal of the Acoustical Society of America*, 71/3, pp. 679-88

TERHARDT, Ernst, Gerhard STOLL, and Manfred SEEWANN (1982b), 'Pitch of Complex Signals According to Virtual Pitch Theory: Tests, Examples, and Predictions', *The Journal of the Acoustical Society of America*, 71/3, pp. 671-8

VAN NOORDEN, Leo Paulus Antonie Servatius (1975), *Temporal Coherence in the Perception of Tone Sequences*, (PhD dissertation, Eindhoven, Druk vam Voorschoten)

Gilberto Bernardes has a multifaceted activity as a saxophonist, new media artist, and researcher in sound and music computing. He performs regularly in Europe and the United States, and has been published in many peer-reviewed scientific journals and international conferences. Gilberto Bernardes holds a PhD in digital media from the University of Oporto (under the auspices of the University of Texas at Austin), a Master of Music from the Conservatory of Amsterdam, a Bachelor of Music from the Polytechnic Institute of Oporto, and a Degree ('Premier Prix') in saxophone and chamber music from the ENM d'Issy-les-Moulineaux (Paris, France). Since 2007 he has been developing research into generative music systems through different projects, including his most current position as a senior researcher at the Sound and Music Computing Group from INESC Technology and Science. Gilberto Bernardes is currently an Invited Adjunct Professor at the Polytechnic Institute of Castelo Branco.