# Commentary on Cook & Fujisawa's "The Psychophysics of Harmony Perception: Harmony is a Three-Tone Phenomenon"

RICHARD PARNCUTT
*University of Graz*

ABSTRACT: Cook & Fujisawa (2006) point out that, contrary to the predictions of psychoacoustic models, the diminished triad is more consonant and prevalent in western tonal music than the augmented. A possible simple explanation is that the diminished triad often functions as an incomplete dominant (major-minor) seventh chord, the most prevalent tetrad in mainstream tonal music.

I agree with many of the points raised in this article (Cook & Fujisawa, 2006) and thank the authors for raising some interesting issues. I agree, for example, that there is more to the perception of a chord than the sum of the perception of the intervals. I also agree that, so far, no psychoacoustical model has succeeded in predicting the relative perceived consonance of common musical chords such as the major, minor, diminished and augmented triads.

Regarding the latter point, Eberlein (1994) investigated the prevalence (frequency of occurrence) of chords and chord progressions in a sample of music from the 18th and 19th centuries: Bach Chorales, Händel trio sonatas, Mozart and Beethoven masses, Mendelssohn motets, and Bach's *Kleine harmonische Labyrinth* BWV 59. Table 1 shows the results for sonorities of three pitch classes. Since the prevalence of musical elements in tonal Western music primarily depends on their consonance (cf. Huron, 1991), the table supports Cook and Fujisawa's point that the diminished triad is more consonant than the augmented.

| Music-theoretical name | Symbol | Example on C | Root position | 1st inversion | 2nd inversion | Total |
|---|---|---|---|---|---|---|
| Major triad | | CEG | 513 | 200 | 68 | 781 |
| Minor triad | m | CEbG | 322 | 157 | 83 | 562 |
| Diminished triad | o | CE$^b$G$^b$ | 25 | 90 | 13 | 128 |
| Suspended triad | sus | CFG | 32 | <7 | <7 | <46 |
| Augmented triad | + | CEG# | <7 | <7 | <7 | <21 |

**Table 1.** Prevalence of chords of three pitch classes in a sample of music from the 18th and 19th centuries according to Eberlein (1994, p. 421). For example, he located 513 root-position major triads in his sample. He did not publish data for chords that occurred fewer than 7 times.

Cook and Fujisawa rightly challenge the claim in Parncutt (1989) that the perceptual dissonance of the augmented triad has cultural rather than sensory origins. I see now that this claim was a cop-out: if the sensory model doesn't work, try another explanation. With that kind of approach, it is impossible to falsify the model.

I think I can now offer a more satisfactory explanation. My explanation is deliberately simple, because parsimonious models tend to be less arbitrary and easier to falsify. I begin with the following three assumptions.

1. The dissonance of a chord type in the western tonal system (such as the major triad or diminished seventh) may be analysed into three main components: roughness, root ambiguity, and unfamiliarity (cf. Terhardt, 1977). Or conversely: consonance is a combination of smoothness, root clarity, and familiarity. These three parameters vary along a scale from

almost completely physiologically determined and culture-independent at one end (roughness) to completely culturally determined at the other (familiarity).

    a. Roughness may be regarded as culture-independent in the sense that it is perceived (in the sense of "picked up") in much the same way in all cultures, but its *aesthetic appraisal* is culture-dependent. For example, the traditional two-part polyphony of Balkan countries such as Bulgaria and Serbia plays with the roughness of major and minor second intervals; and loud rock music is deliberately rough, although the underlying harmonic structures are triadic. In both cases, we may safely assume that all listeners experience the sensation of the roughness, but that that sensation is evaluated differently depending on the musical or social context and the listener's cultural background.

    b. I introduced the term *root ambiguity* in Parncutt (1988, 1989). A chord with a clear root tends to have a clear function in a harmonic progression, so one may assume that chords with clear roots are (or were) preferred by composers of tonal music. Chords with clear roots also fuse better in the sense of Stumpf's (1890) *Verschmelzung*, but I am not aware of any clear reason why composers, performers or listeners should generally prefer or avoid sonorities that fuse. For example, Huron (1991) showed that Bach avoided *intervals* that fuse well (octaves, fifths) if they compromised the independence of the melodic lines, while at the same time favoring *chord* types that fuse well (major, minor, dominant seventh)—presumably for their clarity harmonic function.

    c. Regarding familiarity: I assume that any musical element can become more consonant if it is heard more often, either by an individual or a whole culture (cf. Peretz, Gaudreau, & Bonnel, 1998; Szpunar, Schellenberg, & Pliner, 2004). In this regard it is important to distinguish clearly between consonance and preference: even a sound that is heard so many times that it becomes boring or unpleasant (Berlyne, 1970) may at the same time be perceived as consonant because of its familiarity.

2. The first two aspects of consonance/dissonance, smoothness/roughness and root ambiguity, are related in a simple way to the interval structure of chords.

    a. Chord types that contain the interval of a major or minor second tend to evoke more roughness than others.

    b. Chord types that contain the interval (class) of a perfect fifth (or fourth) tend to have clearer roots and harmonic functions than others.

In both cases, the expression "tend to" alludes to the fact that chord types can appear in different, voices, registers, and spacings, as well as timbres—so there is considerable variation in fusion and roughness among different voicings of the same chord class.

3. Of all 19 possible triad types in the chromatic scale (12 pitch class sets of cardinality 3, plus the inversions of those 7 that are not symmetrical; Rahn, 1980), only two contain a perfect fifth (or fourth) and no seconds: the major and minor triads. I made this apparently trivial observation in Parncutt (1988). Perhaps it is not so trivial after all, because it answers basic questions such as "What is so special about the major and minor triads?" and "Why does harmonic tonality have two modes". I know of no more plausible answers to these questions that are at once simple (falsifiable), grounded in direct observation, and part of a broader theory. Incidentally, only two other triad types contain no seconds: the diminished and augmented triads.

These three assumptions can explain firstly why the major and minor triads, taken together, are more consonant (and prevalent) than the diminished and augmented triads, again taken together. Beyond that, they explain why the major triad is more consonant than the minor: it has a clearer root. This also explains for example why the Renaissance-Baroque compositional practice of *tierce de Picardie* involves a major triad following a minor passage – and not the other way around. The pitch salience profiles of Parncutt (1988) suggest that the third of a minor triad competes with the (conventional) root for the status of (perceptual) root; similarly, the key profiles of Krumhansl & Kessler (1982) suggest that the third scale member of a minor key competes with the (conventional) tonic for the status of (perceptual) tonic.

This line of argument cannot, however, directly explain why the diminished triad is considered more consonant than the augmented, both in music theory and in listening tests (Roberts, 1986, cited by Cook & Fujisawa). The diminished triad is more dissonant than the augmented in terms of roughness,

and less dissonant in terms of root ambiguity. Perhaps root ambiguity makes a bigger contribution than roughness to the difference in overall dissonance and prevalence between these two chords. Similarly, the difference between the consonance of major and minor triads appears to be primarily due to the clarity of the root.

Cook and Fujisawa argue that the consonance of a chord type depends only on the consonance of the chord as a whole, and not (or not directly) on the individual intervals of which it is composed. This is consistent with my approach in Parncutt (1988, 1989), as well as for example Hutchinson and Knopoff (1978), whose algorithm can be applied to any sonority of any number of notes or partials. But the above arguments suggest that the situation may be more complex: the consonance of a chord type may, in fact, depend *directly* on *both* the chord as a whole *and* the individual intervals. Consider the following points:

1.  It is common in musical performance to bring out one of the tones in a musical chord – the "melody" (regardless of whether it is placed in the soprano, bass or an inner voice). This can considerably influence the chord's perceptual roughness. Terhardt (1974) demonstrated that the magnitude of the perceived roughness of an amplitude-modulated waveform is approximately proportional to the square of the degree of modulation. Consider for example a harmonic dyad of pure tones: when the amplitude of one tone is reduced to half that of the other (-6 dB), the perceived roughness falls to less than half its previous magnitude. Thus, if one tone in a triad is performed relatively quietly, the roughness of the intervals between that tone and the other tones may become perceptually irrelevant. The roughness of the triad may depend primarily on the roughness of the interval between the other two tones.

2.  By contrast, the ambiguity of the root of a chord is essentially unrelated to the ambiguity of its individual intervals. For example, the augmented triad comprises only major third intervals (ignoring the separate issue of enharmonics), each of whose root corresponds clearly to the lower tone; but the root of the triad is highly ambiguous. And the major-minor (dominant) seventh chord has a very clear root, even though it contains the ambiguous tritone interval. Furthermore, the clarity of the root of a chord appears to be affected relatively little by the relative loudness of its tones, provided all are clearly audible.

The argument of Cook and Fujisawa relies on Meyer's (1956) observation that equally-spaced pitch patterns are intrinsically more dissonant than non-equally spaced patterns. The usual (and as far as I know the only plausible) reason for this is that *symmetry* is associated with *ambiguity*. The whole-tone scale has no tonic because all six tones are equally likely to be the tonic. The diminished and augmented triads have no roots because all tones are equally good root candidates.

The connection between equal spacing and consonance is thus *indirect* and mediated by cognition and culture. Consider a scenario in which the augmented triad is not yet recognized as such (e.g. by theorists), but occasionally finds its way accidentally into musical works, as a result of the voice-leading. Since the inherent ambiguity of this pitch combination is not immediately evident, the perception of its consonance or dissonance cannot, initially, be affected by its ambiguity. Its ambiguity only becomes clear, and has a chance to affect its perceived consonance or dissonance and its musical prevalence, once the element finds its way into an accepted musical vocabulary. This generally follows an initial period of experimentation (trial and error). Thus, the association between equal pitch spacing and dissonance cannot be regarded as "psychoacoustic" – at least, not in the way that the dependence of the pitch, loudness, roughness, timbre and perceived duration of a complex sound on its physical properties is "psychoacoustic".

I find Cook and Fujisawa's treatment problematical in the following additional respects.

1.  The Gaussian curve specified by Equation 1 is unconvincing. Equal spacing is surely perceived categorically - it is not a continuous parameter. A pattern is either equally spaced or it is not. There is no half-way stage in between.

2.  While it is true that augmented triads, diminished seventh chords, and whole-tone scales are dissonant by comparison to other, more common triads, tetrads and scales (respectively), the main reason for this difference may not be symmetry per se, but the simple fact that these pitch configurations do not include the root-support interval of a perfect fifth or fourth. That explains directly and simply why their root and harmonic function (or in the case of the scale, their tonic and tonal properties) is unclear.

3.  The theory of Cook and Fujisawa is similar to the mathematical music theories of Balzano (1980), Mazzola (1990) and Noll (2004) in that mysterious emergent properties of mathematical models are assumed to override the everyday ("psychoacoustic") experience of the roughness and fusion of music intervals and chords. Although these mathematical theories impress at first glance, they lack clarity and parsimony. They are difficult to falsify and therefore not scientific in that sense. They are incomplete, because they fail to make a concrete *historical* link between the described mathematical properties and real music. There is a need to compare the predictions of such models, based on the mathematical possibilities inherent in musical scales, with statistical distributions of real musical structures from different historical periods and geographical regions.

Since the entire mathematical treatment of Cook and Fujisawa appears to be based on a problematic assumption, I will not consider the details here. Instead, I will try to develop an alternative model of the consonance of any chord playable in the chromatic scale that answers their questions.

Clearly, there is something wrong with my model of root ambiguity: both the octave-generalised version (Parncutt, 1988) and the non-octave-generalised version (Parncutt, 1989) predict incorrectly that the root of an augmented triad is less ambiguous than the root of a diminished triad (i.e. that the augmented triad is more consonant in this respect). Since both these algorithms are in fact quantifying *pitch* ambiguity and assuming that this equates to *root* ambiguity, a possible explanation is that pitch and root ambiguity are not quantitatively equivalent. But how else can root ambiguity be predicted for any subset of the chromatic scale? I cannot think of a solution that does not sacrifice other aspects of the generality of existing algorithms or introduce additional complexity and, therefore, arbitrariness.

| Music-theoretical name | Symbol | Example on C | Root position | 1st inversion | 2nd inversion | 3rd inversion | Total |
|---|---|---|---|---|---|---|---|
| Major-minor 7th | 7, Mm7 | CEGB$^b$ | 140 | 73 | 7 | 37 | 257 |
| Minor 7th | m7 | CE$^b$GB$^b$ | 30 | 34 | <7 | 7 | <78 |
| Half-diminished 7th | Ø7 | CE$^b$G$^b$B$^b$ | 10 | 36 | 10 | <7 | <63 |
| Diminished 7th | o7 | CE$^b$G$^b$B$^{bb}$ | - | - | - | - | 55 |
| 7th, suspended 4th | 7sus4 | CFGB$^b$ | 8 | <7 | <7 | <7 | <29 |
| Major 7th | M7 | CEGB | <7 | <7 | <7 | <7 | <28 |

**Table 2.** Prevalence of chords of four pitch classes in a sample of music from the 18th and 19th centuries (Eberlein, 1994, p. 421). Eberlein did not publish data for chords that occurred less than 7 times.

An improved quantitative model of root ambiguity might consider the specific roles played by the chords in question in real musical contexts. The prevalence of the diminished triad in tonal music may be due to its status as a subset of the major-minor seventh chord (e.g. B-D-F followed by C-E implies the chord progression from G$^7$ to C). Its prevalence may therefore be related to the prevalence of the major-minor seventh chord. According to Eberlein (1994), the major-minor or dominant seventh chord is the most prevalent seventh chord in music of the 18th and 19th centuries (see Table 1). This is consistent with the observation that the calculated root ambiguity of the major-minor seventh chord (according to both Parncutt 1988 and Parncutt 1989) is lower than that of any other seventh chord or set of 4 pitch classes.

Incidentally, Eberlein's observation breaks down for much international Anglo-American popular music in recent decades, which deliberately avoids both dominant seventh chords and diminished triads. However, this does not contradict the theory of root ambiguity, because there is another reason for avoiding these elements. Recent popular music may simply have reacted to and avoided the tonal cliché of a tritone within a dominant seventh resolving to a third within the tonic triad – the "classic turn of phrase" investigated by Gjerdingen (1988) – just as post-medieval music avoided the medieval cliché of parallel fifths. Style can, and often does, contradict psychoacoustics.

If it is true that the diminished triad is prevalent because it often represents the dominant seventh chord, the consonance of the diminished triad relative to the augmented may thus be largely due to familiarity —which is presumably what I meant in Parncutt (1989) by the claim that the effect is cultural rather than sensory in origin. In a nutshell: The major-minor seventh chord is prevalent

because it has a clear root (in spite of the dissonance of the tritone between its third and seventh), and the diminished triad is prevalent because it is can represent that chord (e.g. in a three-voice context).

A two-step explanation of this kind would be hard to encapsulate in a general quantitative model. Such a model might first evaluate the consonance of all possible sonorities playable in the chromatic scale, and on that basis estimate their prevalence in tonal music. In a separate stage, it would list all the chords of which any given chord may be a subset (using pitch-class set theory), and adjust the predicted consonance of each subordinate chord according to the predicted prevalence of its various superordinate chords. To fit the predictions of such a model to empirical data (such as consonance ratings of sonorities presented in isolation, or statistical counts of the prevalence of sonorities in a representative databank of tonal music), one would have to adjust a number of free parameters—a greater number than I adjusted in existing algorithms of root ambiguity. The model may be too arbitrary and not parsimonious enough to allow for the possibility of falsification (Ockham's razor).

Before closing I would like to remind readers that the "root support weights" published in Parncutt (1988) were later corrected without changing any other part of the algorithm. In Parncutt (1988), I assumed that the "root-support intervals" include not only the five octave-generalized intervals corresponding to the first 10 harmonics - P1 (unison or octave), P5 (perfect fifth), M3 (major third), m7 and M2—but also the m3. I later realized that my argument for including the m3 was flawed and subsequently left it out. It is certainly not necessary to include the m3 among the root support intervals in order to explain the root and consonance of the minor triad or any other chord. The m3 was omitted from the root support intervals in all calculations that I published in all later papers. I also adjusted the weights of the root-support intervals in the algorithm to P1=10, P5=5, M3=3, m7=2 and M2=1.

A final word about that minor triad. Generations of music theorists grappled unsuccessfully with the problem of explaining its root and its consonance—among them Rameau, Hauptmann, Oettingen, Schoenberg and Hindemith. As far as I know, my 1988 adaptation of Terhardt's (1982) model has finally solved this long-standing problem. At the same time, it explains the root and consonance of any simultaneity of tones from the chromatic scale. It also explains the ambiguity of the root, by predicting not only the main root candidate but also other possible candidates, ranked by salience. I would be grateful for more feedback from the music theory community about this claim—not only in the context of post-diatonic music (Väisälä, 2002), but also regular diatonic styles and associated approaches such as neo-Riemannian theory.

## REFERENCES

Balzano, G. (1980). The group-theoretic description of 12-fold and microtonal pitch systems. *Computer Music Journal,* Vol. 40*,* pp. 66–84.

Berlyne, D. E. (1970). Novelty, complexity, and hedonic value. *Perception & Psychophysics,* Vol. 8, pp. 279–286.

Cook, N., & Fujisawa, T. X. (2006). The psychophysics of harmony perception: Harmony is a three-tone phenomenon. *Empirical Musicology Review*, Vol. 1, No. 2, pp. 106-126.

Eberlein, R. (1994). Die Entstehung der tonalen Klangsyntax. Frankfurt: Lang.

Gjerdingen, R. O. (1988). *A Classic Turn of Phrase: Music and the Psychology of Convention.* Philadelphia: University of Pennsylvania Press.

Huron, D. (1991). Tonal consonance versus tonal fusion in polyphonic sonorities. *Music Perception*, Vol. 9*,* pp. 135-154.

Hutchinson, W. & Knopoff, L. (1978). The acoustic component of Western consonance. *Interface*, Vol. 7, pp. 1-29.

Krumhansl, C. L., & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review*, Vol. 89, pp. 334-368.

Mazzola, G. (1990). *Geometrie der Töne: Elemente der mathematischen Musiktheorie.* Basel: Birkhäuser.

Meyer, L.B. (1956). *Emotion and Meaning in Music*. Chicago: University of Chicago Press.

Noll, T. (2004). The topos of triads. In H. Fripertinger & L. Reich (Eds.), *Colloquium on Mathematical Music Theory* (Grazer Mathematische Berichte Nr. 247) (pp. 1-26). Graz: Institut für Mathematik der Universität Graz.

Parncutt, R. (1988). Revision of Terhardt's psychoacoustical model of the root(s) of a musical chord. *Music Perception*, Vol. 6, pp. 65-94.

Parncutt, R. (1989). *Harmony: A psychoacoustical approach*. Berlin: Springer-Verlag.

Peretz, I., Gaudreau, D., & Bonnel, A. M. (1998). Exposure effects on music preferences and recognition. *Memory & Cognition*, Vol. 15, pp. 379–388.

Rahn, J. (1980). *Basic Atonal Theory*. New York: Longmans.

Roberts, L. (1986). Consonance judgments of musical chords by musicians and untrained listeners. *Acustica,* Vol. 62, pp. 163-171.

Stumpf, C. (1890). *Tonpsychologie* (Vol. 2). Leipzig: Hirzel.

Szpunar, K. K., Schellenberg, E. G., & Pliner, P. (2004). Liking and memory for musical stimuli as a function of exposure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol. 30, pp. 370–381.

Terhardt, E. (1974). On the perception of periodic sound fluctuations (roughness). *Acustica*, Vol. 30, pp. 201-213.

Terhardt, E. (1977). The two-component theory of musical consonance. In E.F. Evans and E.P. Wilson (eds.), *Psychophysics and Physiology of Hearing* (pp. 381-390). London: Academic Press.

Terhardt, E. (1982). Die psychoakustischen Grundlagen der musikalischen Akkordgrundtöne und deren algorithmische Bestimmung. In C. Dahlhaus & M. Krause (Hrsg.), *Tiefenstruktur der Musik* (pp. 23-50). Berlin: Technical University of Berlin.

Väisälä, O. (2002). Prolongation of harmonies related to the harmonic series in early post-tonal music. *Journal of Music Theory,* Vol. 46, pp. 207-263.