



## Book Review: Robert Cogan, *New Images of Musical Sound*, Harvard University Press, 1984

Stephen W. Smoliar



KEYWORDS: Fourier analysis, music analysis, phonology, sonology, auditory perception

ABSTRACT: While Robert Cogan's *New Images of Musical Sound* is now over ten years old, the applicability of Fourier analysis as a basis for music theory is still a relevant issue. This review attempts to put Cogan's work into both a technical perspective, regarding what sound analysis technology now supports, and a theoretical one, regarding his attempt to build a theory on the foundations of phonology. From a technical point of view, it is now far easier to approach analysis as Cogan has done; but his attempt to build a theory suffers from some significant.

---

### The Need to Look at Sounds

[1] This book is now over ten years old. However, a recent discussion thread on mto-talk was addressing the applicability of Fourier analysis as a basis for music theory; and, in the course of this discussion, David Lewin was kind enough to observe that Robert Cogan had already discussed some of these issues. It therefore seemed appropriate to return to this book and see how it has withstood the past decade.

[2] Cogan's primary interest is in music analysis. Given how much disagreement there has been over just what music analysis is all about, he clarifies his own position in his final chapter: "To analyze is to create a map or model—a model that reveals certain functions and relationships" (page 153). He then develops this point along a line very similar to one which John Roeder recently presented in *Music Theory Online*.<sup>(1)</sup> "A model itself can be verbal, numerical, or graphic, and the preceding chapters have employed (to varying degrees) all of these means: commentaries, tables of oppositions, and spectral photos" (Cogan, page 153). Roeder's own interpretation of the graphic had more to do with the use of diagrams to explicate mathematical relationships;<sup>(2)</sup> and, of course, just about every form of music notation is also graphic. However, Cogan's approach is far more direct: He is interested in graphics to the extent that they satisfy his need to *look at sounds*, and his thesis is that this need may be best satisfied by spectrographic traces of those sounds. In this way what the eye sees in such traces can supplement what the ear hears, perhaps even informing the mind of structural details which are not immediately apparent in the course of listening.

[3] Before examining the specifics of this approach, it is worth reviewing some of its general virtues. Most important is what may be called Cogan's attempt to "conquer time." Sound cannot exist without the passage of time; but, during that passage,

the sound goes as soon as it comes, so to speak. An instant cannot be scrutinized because, once scrutiny begins, the instant is gone. Cogan's images, on the other hand, are *traces* which remain in the present long after the sound has faded into the past; and, unlike the sounds themselves, those traces *can* be scrutinized in as much or as little time as the mind chooses to allocate. Music notation, of course, has the same advantage of timelessness; but notation is, at best, a *prescription* for sound. Cogan has attempted, for purposes of *description*, to capture *the sound itself*.

[4] Another advantage to Cogan's approach is its scalability. By suitably compressing the time scale, one can take in the entirety of any sound event in a single glance. Of course if that event happens to be all of *Goetterdaemmerung*, that single glance is not likely to take in very much detail; but one can then adjust the time scale in order to examine greater detail. In other words if the data are appropriately captured, the observer can control what that single glance encompasses, moving between coarse and fine detail at will. The variable time scale provides an *implicitly hierarchical* view of the sounds of any musical experience, which gets away entirely from the symbolic representations of hierarchy found in the approaches of Heinrich Schenker<sup>(3)</sup> and Eugene Narmour.<sup>(4)</sup>

[5] However, spectrographic traces are only one of many ways in which sound events may be represented visually.<sup>(5)</sup> There is also the waveform itself: a display of how, for example, a loudspeaker cone physically vibrates with the passage of time. This is again a display with the advantage of scalability; but, in this case, we have to be more careful about the scales at which we choose to examine the signal. When we go down to the millisecond level, we can see the actual periodic waveforms associated with isolated pitches. However, the *shape* of such a waveform depends not only on its *harmonic* content (as revealed by a spectrogram) but also on the *phase* associated with each of the component frequencies. The problem is that significant differences in phase do not necessarily imply differences in what is heard.<sup>(6)</sup> **Figure 1** illustrates two waveforms, each of which is represented by a single cycle. Both waveforms have the same harmonic content in identical proportions; but, in the second waveform, the first and second harmonics are cosines, rather than sines, which means they have been phase-shifted by 90 degrees. However, in spite of the obvious differences in appearances, these waveforms are indistinguishable to the ear.

[6] On the other hand if we view these waveforms at a *macroscopic*, rather than *microscopic*, level, they have at least the potential of being more informative. **Figure 2**, for example, is the entirety of the Aloys Kontarsky recording of Karlheinz Stockhausen's "Klavierstueck III".<sup>(7)</sup> This display tells us nothing about notes or the serial structure of the pitches, but it *does* show how those notes are grouped into *gestures* and how the intensities of those gestures are modulated.

[7] Before examining *any* such visual approach in greater detail, however, we should also be sober enough to recognize that *biological* support is not encouraging. One thing we know for certain is that the path from ear to brain is decidedly different from that from eye to brain.<sup>(8)</sup> Thus, if the biological substrates differ, it is unlikely that principles which dictate how we identify objects and structures in what we see are necessarily going to carry over into what we hear. Furthermore, much of this distinction has to do with the necessity to account for time in auditory perception. It is all very well and good to synthesize visual traces which "freeze" the passage of time; but once the stimuli are "frozen," they can no longer be auditory. If time "stops" then so do the stimuli; and any attempt to abstract away from the passage of time runs the risk of also abstracting away certain attributes and relations which may be most critical to how those stimuli are perceived and interpreted.

### Cogan's Approach

[8] Probably the element which has changed the most in the ten years since Cogan published these results has been the supporting technology. Cogan's data were collected during the years 1980 and 1981 in the Sonic Analysis Laboratory at the New England Conservatory. The very first sentence in the book acknowledges the support of Dale Teaney and Charles Potter from the IBM Watson Research Center: "Without their professional and personal initiatives, the process would have become available to musicians much later than it did" (p. v).

[9] Given the extent to which computers have become part of our day-to-day lives, it is hard to believe that, when these data were being collected, personal computing did not yet exist. Much of what we now take for granted could not even be imagined in 1980, making it somewhat difficult to infer from the description in Appendix A just what instrumentation Cogan actually used to arrive at the spectrum photos which lie at the heart of this book. He *does* tell us that the "spectrum

analyzer was a thirty-three- millisecond fast Fourier transform instrument, capable of analyzing sounds in five contiguous octave registers simultaneously” (page 155); so we can conclude that at least *some* of the equipment was digital. On the other hand the images appear as if they were photographed from a rather conventional (and probably temperamental) analog oscilloscope; and there is no doubt that he had to use photography (rather than, for example, laser printing) to capture those images (not to mention physical acts of cutting and pasting his photographs, rather than composing his images with software assistance). However, there are also references to dynamic controls which could be adjusted, with little reference to what exactly is being controlled.

[10] One thing is certain: Anyone interested in undertaking a similar project today is going to have a far easier time of it. These days it is pretty difficult to find a personal computer which *lacks* some form of audio input, not to mention direct capture of data from audio compact disc recordings. It is quite likely that Cogan could now carry all the examples from his book around in a laptop computer, examining and listening to his data while on a flight surrounded by other users who are buried deep in their spreadsheets and games (all of whom, collectively, are probably radiating enough frequency to give the pilot a whole new patch of gray hairs). Put another way, this is no longer a big deal, which means that, putting a more positive slope on the situation, it is now the sort of thing we can expect any resourceful undergraduate to do.

[11] To be a bit more fair, all this means is that *collecting* data is no big deal. The “real deal” is what happens next. Therefore, I would like to be relatively brief in reviewing the data which Cogan actually collected and focus more attention on how he then undertook to develop a theory from them.

### The Examples

[12] Part I of the book is essentially an exposition of seventeen spectrum photos. These are collected into four chapters: Voices, Instruments, Large Mixed Ensembles, and Electronic and Tape Music. Each chapter, in turn, attempts to explore considerable variety in its subject matter. Thus, the Voices chapter covers Gregorian chant, Tibetan Tantric chant, Billie Holiday singing “Strange Fruit,” and Gyorgy Ligeti’s “Lux Aeterna.” The instrumental examples include a Balinese gamelan, a Ludwig van Beethoven piano sonata recorded on both a forte-piano and a modern instrument, the second of Igor Stravinsky’s “Three Pieces for String Quartet,” the latter two of Anton Webern’s Opus 7 (“Four Pieces for Violin and Piano”), and the third etude from Elliott Carter’s “Eight Etudes and a Fantasy” (the only example of winds). The mixed ensembles include both instruments and voices. The first two examples compare the “Confutatis” from Wolfgang Amadeus Mozart’s *Requiem* with the “Tibi Omnes Angeli” from Hector Berlioz’ *Te Deum*. This is followed by the first of Claude Debussy’s orchestral “Nocturnes,” the brief orchestral interlude which follows Marie’s murder in Alban Berg’s *Wozzeck*, and Edgard Varese’s *Hyperprism*. In contrast the final chapter is rather sparse in its examples: the Introduction to Milton Babbitt’s “Ensembles for Synthesizer,” the “Fall” movement from Jean-Claude Risset’s *Little Boy* suite, and Cogan’s own “No Attack of Organic Metals.” Nevertheless, the overall variety of the entire collection is quite satisfying; and it is gratifying to see that Cogan deliberately avoided concentrating on the music of Dead White European Males.

### The Theory

[13] What is less satisfying about Part I of the book is that it is primarily anecdotal. As one proceeds through the images, one gets the feeling that Cogan is saying, “Here is something interesting. Here is something else interesting. Here is yet another interesting thing.” After a while, even the most generous reader is likely to erupt: “YES! I agree! There *are* interesting things here! Do you also have an interesting *theory*?”

[14] Cogan attempts to confront this question with Part II of his book. This is a decidedly shorter part of the entire volume; and, unfortunately, it is also noticeably weaker. Some of its weaknesses may be more apparent because of what we have learned over the last ten years; but, even in his own time, Cogan was failing to do justice to his subject matter in several significant ways.

[15] In order to develop his theory, Cogan turns to *phonology*, that division of linguistics which is concerned with the *sounds* of language, as opposed to either syntax or semantics. This was not, even in its own time, a particularly new idea. Gottfried-Michael Koenig had organized an Institute for Sonology at the University of Utrecht in a deliberate attempt to generalize

phonological theory to encompass other sources of sound (such as experiments in electronic and computer music which particularly interested Koenig); and some of the earliest research concerned with developing a sonological theory was undertaken by Otto Laske.<sup>(9)</sup> It is therefore unfortunate that Cogan gives no indication of either Koenig or Laske, either the theory or practice of their work, or even the *idea* of generalizing from phonology to sonology. Instead, his primary foundations seem to rest on the work of Roman Jakobson<sup>(10)</sup> and N. S. Trubetzkoy.<sup>(11)</sup> There seems to be little acknowledgment that there might be any *other* foundations, such as those which were pursued by Morris Halle and Noam Chomsky;<sup>(12)</sup> so the reader is left with the uneasy feeling that Cogan decided to pursue this particular approach because it looked good at the time.

[16] What interested Cogan most was Jakobson's desire to describe sounds in terms of *oppositions*. This amounts to describing the quality of a sound in terms of where it is situated between two contrasting extremes. For example, the grave/acute opposition distinguishes between concentration of sounds in low and high frequencies, respectively. On the other hand the centered/extreme opposition addresses whether there is frequency activity at the extremities of the spectral range or only in some more limited middle range. Cogan presents thirteen of these oppositions; and, while he is very up front about the fact that this list may not be complete, the idea of building such a list in the first place lies at the heart of his theory.

[17] This approach is not unfamiliar in the technology of signal analysis. However, what Jakobson called *oppositions* are now more commonly known as *features*; and, when several of these are collected together, the results are called vectors in a *feature space*.<sup>(13)</sup> Feature vectors have become invaluable tools for the description and recognition of both visual and auditory signals; but, like most other tools, they religiously follow the GIGO (Garbage-In-Garbage-Out) Principle. If they are not used properly, they are likely to yield results which are, at best, questionable. Therefore, it is important to review some questions which need to be asked before feature vectors are invoked as a descriptive tool.

[18] Perhaps the most important question to be asked of any feature is: *Can it be effectively calculated?* This question has been deliberately formulated using the language of Alonzo Church;<sup>(14)</sup> so a positive answer implies that there is some *effective algorithm* which may be applied to the input, whether from spectra or waveforms, which will assign a value for that feature in a reproducible manner. Cogan is quick to point out that his feature values are context-dependent; but that does not preclude their being effectively computed. Computation may just have to look at a broader span of time than, say, thirty-three milliseconds. However, while Cogan is certainly *trying* to be specific in describing his features, effective computation appears to be beyond his scope.

[19] The opposite question is: *What can be reconstructed from a feature vector?* One of the interesting things about the spectrum is that it can be used to reconstruct the original sound.<sup>(15)</sup> Given a feature vector, can it be used to construct *any* sound? The context-dependent nature of Cogan's approach to description definitely works against him on this count. However, even if his value assignments were *less* context-dependent, it is unclear that one could ever look at a feature vector and hear anything remotely relevant in "the mind's ear."

[20] This leads to yet another question: *How do we compare feature vectors for similarities?* This is a familiar problem in visual processing. While any color can be easily described as a weighted sum of red, green, and blue sources of light, colors which are visually similar do not always have similar weight contributions. Color similarity is usually better represented in terms of a different set of descriptors, commonly called luminance, hue, and saturation.<sup>(16)</sup> The two sets of descriptive vectors are equally effective in reconstructing a color, but the second representation facilitates identifying colors which are *perceptually similar*. Thus, even if we have an effective set of features for describing sounds, until we know how to relate those features to auditory perception, we have little more than a mathematical abstraction.

[21] The moral of the story is that the path Cogan has chosen is probably heading in the right direction, but he has not yet properly equipped himself for the journey. Fortunately, our understanding of auditory perception has come a long way since Cogan's book appeared.<sup>(17)</sup> If Cogan is no longer interested in the trip, others can look where his finger is pointing and set off on their own. Thus, the weakness of the theoretical portion of this book should not be seen as a condemnation of Cogan's approach to collecting data but as an incentive to apply our increased knowledge of both signal analysis and perception to go forth and do a better job.

## Conclusions

[22] Is the trip going to be worth making, bearing in mind, for example, the biological conflict between what we hear and what we see? If the purpose of the trip is to try to reduce all that is auditory to the visual, then, most likely, the trip will be doomed to failure. However, as Cogan observed, images are but one of many approaches to description; and no one approach will ever do all the work.<sup>(18)</sup> So we should not expect images of either spectra or waveforms to yield *all* the secrets of any musical experiences. This should not be the purpose of the trip. Instead, one should undertake the trip to learn, in more specific ways, what these image data both *can* and *cannot* tell us. If we undertake this task seriously, we are likely to find that those data can, indeed, tell us things which cannot be readily accommodated, if at all, by other modes of description. Such a discovery will leave us better equipped than ever for future analyses of music and a firmer sense of the capabilities of music theory.

---

**Stephen W. Smoliar**  
**Institute of Systems Science**  
**National University of Singapore**  
**Heng Mui Keng Terrace**  
**Kent Ridge 0511**  
**SINGAPORE**  
[smoliar@iss.nus.sg](mailto:smoliar@iss.nus.sg)

---

## Footnotes

1. Roeder, J. 1993. "Toward a Semiotic Evaluation of Music Analysis," *Music Theory Online* 0.5: 4.  
[Return to text](#)
2. Ibid, 8.  
[Return to text](#)
3. Schenker, H. 1956. *Der Freie Satz*, O. Jonas, editor, Vienna: Universal Edition.  
[Return to text](#)
4. Narmour, E. 1977. *Beyond Schenkerism: The Need for Alternatives in Music Analysis*. Chicago: The University of Chicago Press.  
[Return to text](#)
5. Aigrain, P., et al. 1995. Representation-Based User Interfaces for the Audiovisual Library of the Year 2000, *Proceedings: Multimedia Computing and Networking 1995*, A. A. Rodriguez and J. Maitan, editors, SPIE, 35–45.  
[Return to text](#)
6. Risset, J.-C. 1991. Timbre Analysis by Synthesis: Representations, Imitations, and Variants for Musical Composition, *Representations of Musical Signals*, G. DePoli, A. Piccialli, and C. Roads, editors, Cambridge: The MIT Press, 7–43.  
[Return to text](#)
7. The sound was digitized from the vinyl Columbia recording, 32 31 0008; these recording sessions were supervised by Stockhausen.  
[Return to text](#)
8. Gibson, J. J. 1983. *The Senses Considered as Perceptual Systems*. Westport: Greenwood Press.  
[Return to text](#)
9. Laske, O. 1972. "On Problems of a Performance Model for Music" (Technical Report, Institute of Sonology, Utrecht State University): 29. For a critical review of this work, see Smoliar, S. W. 1976. "Music Programs: An Approach to Music Theory

Through Computational Linguistics,” *Journal of Music Theory* 20.1: 105–131.

[Return to text](#)

10. Jakobson, R., and Waugh, L. 1979. *The Sound Shape of Language*. Bloomington: Indiana University Press.

[Return to text](#)

11. Trubetzkoy, N. S. 1969. *Principles of Phonology*, C. A. M. Baltaxe, translator, Berkeley: University of California Press.

[Return to text](#)

12. Chomsky, N., and Halle, M. 1968. *The Sound Pattern of English*. New York: Harper & Row.

[Return to text](#)

13. Gianotti, C. 1993. Analysis of Economic and Business Information, *Handbook of Pattern Recognition and Computer Vision*, C. H. Chen, L. F. Pau, and S. Wang, editors, Singapore: World Scientific, 569–594.

[Return to text](#)

14. Church A. 1965. An Unsolvable Problem of Elementary Number Theory, *The Undecidable: Basic Papers On Undecidable Propositions, Unsolvability Problems And Computable Functions*, M. Davis, editor, Hewlett: Raven Press, 88–107.

[Return to text](#)

15. Butler, D. 1992. *The Musician’s Guide to Perception and Cognition*. New York: Schirmer: 208.

[Return to text](#)

16. Luong, Q.-T. 1993. Color in Computer Vision, *Handbook of Pattern Recognition and Computer Vision*, C. H. Chen, L. F. Pau, and S. Wang, editors, Singapore: World Scientific, 311–368.

[Return to text](#)

17. One may gain a good appreciation of how long we have come from Butler, op. cit.

[Return to text](#)

18. Smoliar, S. 1994. “Comment on John Roeder’s Article,” *Music Theory Online* 0.6: 7.

[Return to text](#)

---

## Copyright Statement

**Copyright © 1995 by the Society for Music Theory. All rights reserved.**

[1] Copyrights for individual items published in *Music Theory Online (MTO)* are held by their authors. Items appearing in *MTO* may be saved and stored in electronic or paper form, and may be shared among individuals for purposes of scholarly research or discussion, but may *not* be republished in any form, electronic or print, without prior, written permission from the author(s), and advance notification of the editors of *MTO*.

[2] Any redistributed form of items published in *MTO* must include the following information in a form appropriate to the medium in which the items are to appear:

This item appeared in *Music Theory Online* in [VOLUME #, ISSUE #] on [DAY/MONTH/YEAR]. It was authored by [FULL NAME, EMAIL ADDRESS], with whose written permission it is reprinted here.

[3] Libraries may archive issues of *MTO* in electronic or paper form for public access so long as each issue is stored in its entirety, and no access fee is charged. Exceptions to these requirements must be approved in writing by the editors of *MTO*, who will act in accordance with the decisions of the Society for Music Theory.

This document and all portions thereof are protected by U.S. and international copyright laws. Material contained herein may

be copied and/or distributed for research purposes only.

Prepared by Cara Stroud, Editorial Assistant