



AALBORG UNIVERSITY
DENMARK

Aalborg Universitet

Book review: An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics by Alexander Lerch

Sturm, Bob L.

Published in:
Computer Music Journal

Publication date:
2013

Document Version
Early version, also known as pre-print

[Link to publication from Aalborg University](#)

Citation for published version (APA):

Sturm, B. L. (2013). Book review: An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics by Alexander Lerch. *Computer Music Journal*, 37(4).

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy

If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

This review will be printed in: *Computer Music Journal*, vol. 37, no. 4, Winter 2013.

An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics by Alexander Lerch, October 2012, Wiley-IEEE Press.
ISBN: 978-1-118-26682-3, Hardcover, 272 pages, 503 references. List price \$125.00

Bob L. Sturm

Audio Analysis Lab, AD:MT, Aalborg University Copenhagen, A.C. Meyers Vænge 15, DK-2450 Copenhagen SV, Denmark
E-mail: <bst@create.aau.dk>. Web site: <<http://imi.aau.dk/~bst>>

This book aims to provide a unified and approachable introductory text for a graduate course in analyzing audio music signals. Indeed, it is a result of the author's experience in teaching such a course at Technical University of Berlin, where the only prerequisite is a basic knowledge of digital signal processing. Along with four appendices, it has ten chapters, including an introduction and a review of fundamental concepts in digital signal processing (e.g., sampling), analysis (e.g., Fourier transform), probability theory (e.g. density), and perception (e.g., auditory filterbanks). The main portion of the book, Chapters 3 – 6, presents dozens of different features that have been used in analyzing music audio signals: zero-crossing rate, spectral shape, Mel-frequency cepstral coefficients, and linear prediction coefficients (chapter 3); envelopes and energy features (chapter 4); frequency, pitch, and chroma features (chapter 5); and onsets, tempo, and beat histogram features (chapter 6). The remaining four chapters cover particular applications of audio analysis: dynamic time warping for signal comparison (chapter 7); music similarity and instrument recognition (chapter 8); fingerprinting (chapter 9); and the analysis of musical performance (chapter 10). There are also four appendices covering properties of convolution (A), the Fourier transform and windowing (B), principal component analysis (C), and a review of software for audio analysis (D). The book includes MATLAB code available for free from here: <http://www.audiocontentanalysis.org/>.

I am extremely interested in this text since my department will soon offer such a course. I attempted to read each chapter in two frames of mind: as a lecturer to prepare for and accompany lectures, and as a student to prepare for exams. My search for a suitable text, however, goes on. For the lecturer, this text is confusing, and has so many errors and oversimplifications that too much class time is needed to address them. For the student, the text is very expensive, confusing, and filled with forward references that frustrate reading. While including MATLAB code with a textbook is an excellent idea, and encourages reproducible research [1], the code is not immediately understandable, and none of it readily reproduces any figure in the text. Below, I provide a few illustrations of these criticisms.

Chapter 2, "Fundamentals", presents quantization in Section 2.1.3.2, gives the calculation of the expected quantization error of uniform quantization, and shows its dependence on the probability distribution function (PDF) of the

quantized signal. A student, having no experience with probability theory, will be utterly baffled here until arriving at Section 2.1.4.1 where the PDF is described. However, here the lecturer reads, "The abscissa of a PDF plot represents all possible amplitude values of the signal x and their probability is plotted on the ordinate." Figure 2.5 shows that for a sinewave – which I assume means the sine of a random variable uniformly distributed in $[0, 2\pi]$ – the PDF exceeds 1. Given the text's description of the y-value being a probability, this violates an axiom of probability theory: probability is a scalar in $[0, 1]$. As other examples, Section 2.2.1.2 discusses zero phase filtering, but without context since there is no discussion of phase delay or group delay in filtering. Section 2.2.3 introduces the discrete Fourier transform, and its short-term implementation. While I like the discussion about the practicalities of implementation, the student will find no discussion on what negative frequencies are, the affect of window shapes and sizes, the use of zero padding, frequency response, the significance of magnitude and phase, invertibility, and so on.

The main contribution of the book is its collection in one source of the myriad of features available for signal analysis. However, most of these features are presented without any reference to music audio content. It illustrates most features with a real signal (a "saxophone signal" of about 28 s duration), and three signals having "prototypical spectral shapes": silence (all zeros), "flat magnitude" ("same amplitude at all bins"), and "single peak" (one non-zero value at some frequency). For instance, Table 3.2 and Fig. 3.3 show the "spectral kurtosis" for all these signals. For all "prototypical" signals, this feature is "undefined." For the saxophone signal, we see a plot of a time-series on top of the waveform. The only accompanying description is, "While during the notes high values can be observed, indicating a very peaked distribution, the spectral kurtosis drops significantly during pauses." This provides no illumination on why this feature is important for *analyzing audio content*, what it means for *analyzing audio content*, and when it should be used for *analyzing audio content*. Furthermore, it is strange to consider throughout the book such a trivial signal as all zeros (silence in an audio signal is typically not all zeros).

Often, the text formally describes many of these features in ways that frustrate interpretation: using unconventional notation, not defining all variables, and having errors. For instance, spectral skewness is defined in the text as follows:

$$v_{ssk}(n) = \frac{2 \sum_{k=0}^{\frac{K}{2}-1} (|X(k, n) - \mu_{|X|}|)^3}{K \sigma_{|X|}^3}$$

where K is the length of the analysis window and DFT, $X(k, n)$ is the k th bin of the DFT of the n th signal segment, and $\mu_{|X|}$ and $\sigma_{|X|}^3$ are *undefined*. The expert will know from where those undefined values come, but will also recognize an error: the limit of the sum should be $\frac{K}{2}$ instead of $\frac{K}{2} - 1$. This error appears consistently throughout the text.

The text also contains many simplifications, some of which are wrong and quite mysterious. For instance, the chapter on the temporal analysis of music contains

the following description of *meter*: “a regular alternation of strong and weak musical elements which are grouped with a length of normally three to seven beats or a length of around 5 s” (p. 122). That is *all* the text says about meter. The final four chapters give brief examples of audio analysis. The presentation of dynamic time warping in Chapter 7 is extremely confusing. Chapter 8 discusses audio classification, but nothing about evaluation. Chapter 9 discusses audio fingerprinting, but how this relates to audio content is never explained. (Audio fingerprinting works with perceptually and musically meaningless anchors in a time-frequency space hashed to form an indexable fingerprint.) Finally, chapter 10 is about music performance analysis, but this appears entirely out of context in a book about audio features.

The free code available from the author's website contains several MATLAB scripts for extracting the features described in the text. The current README.txt (accessed June 4, 2013) provides only a brief description of what to do. A student, new to MATLAB or not, will be confused; and the lecturer will have to work to reproduce some figures in the book. Students find it much more preferable, and conducive to experimentation, to have code that unambiguously accompanies each figure [2,3].

Certainly, with 503 references, this text collects a significant amount of work. Overall, though, the textbook lacks the unification, insight, and big-picture perspective of the field of music audio content analysis expected for such a work. Since the discipline of music audio content analysis appears in some respects far from the scientific maturity required for a textbook [4,5,6], the difficulty is very large for such a book to be something that is not “more of the same.” In the mean time, I must assign other reading.

[1] P. Vandewalle, J. Kovacevic, and M. Vetterli, “Reproducible research in signal processing — what, why, and how,” *IEEE Signal Process. Mag.*, vol. 26, pp. 37–47, May 2009.

[2] S. Mallat, *A Wavelet Tour of Signal Processing*. San Diego, CA: Academic Press, 2nd ed., 1999.

[3] B. L. Sturm, Book review, “A Wavelet Tour of Signal Processing, 2nd Edition, by S. Mallat,” *Computer Music Journal*, vol. 31, no. 3, pp. 83-85, Fall 2007.

[4] J.-J. Aucouturier and E. Bigand, “Seven problems that keep MIR from attracting the interest of cognition and neuroscience”, *Journal of Intelligent Information Systems*, 2013 (in press).

[5] J. Urbano, M. Schedl, and X. Serra, “Evaluation in Music Information Retrieval”, *Journal of Intelligent Information Systems*, 2013 (in press).

[6] B. L. Sturm, “Classification accuracy is not enough: On the evaluation of music genre recognition systems”, *Journal of Intelligent Information Systems*, 2013 (in press).