

Commentary on “Comparative Analysis of Music Recordings from Western and Non-Western traditions by Automatic Tonal Feature Extraction” by Emilia Gómez, and Perfecto Herrera

OLIVIER LARTILLOT[1]

*Finnish Centre of Excellence in Interdisciplinary Music Research,
Department of Music, University of Jyväskylä*

PETRI TOIVIAINEN

*Finnish Centre of Excellence in Interdisciplinary Music Research,
Department of Music, University of Jyväskylä*

TUOMAS EEROLA

*Finnish Centre of Excellence in Interdisciplinary Music Research,
Department of Music, University of Jyväskylä*

ABSTRACT: The article by Gómez and Herrera presents an original methodology, audaciously situated on a challenging junction between computer science, cognitive science and ethnomusicology. We hope expert ethnomusicologists will understand the experimental aspect of such a cross-disciplinary undertaking, and will pardon the potential imperfection in this computational attempt toward cross-cultural understanding. Despite the few shortcomings discussed in this commentary, we think the general methodology described in this paper is of high interest.

Submitted 2008 August 25; accepted 2008 August 26.

KEYWORDS: *comparative analysis, western music, non-western music, tonal*

THE article by Gómez and Herrera presents an original methodology, audaciously situated on a challenging junction between computer science, cognitive science and ethnomusicology. In this respect, it relays the proposal to define a new field of study called “Computational Ethnomusicology” (Tzanetakis et al., 2007). We might include in this picture a few works in computational analyses and modeling of non-Western styles of music, and their application to Music Information Retrieval research, published over the last few years and dealing with topics such as pitch scale induction from recordings of African music (Moelants et al., 2006), transcription of carnatic music (Krishnaswamy, 2004) and classification of Indian ragas based on pitch-class distributions (with or without transcription) (Chordia & Rae, 2007).

Dealing with musicological issues, and ethnomusicological issues in particular, is not a simple subject for computer scientists. We hope expert ethnomusicologists will understand the experimental aspect of such a cross-disciplinary undertaking, and will pardon the potential imperfection in this computational attempt toward cross-cultural understanding. The argumentation of the paper might betray for some readers some residual of Eurocentrism. This is apparent, for example, in the way that other cultures are compared to a Western standard, evoking also a description of African culture as an “approximation of the Western tempered scale”, and in citing the notion of “advanced musical systems”. However, we would like to emphasize that the authors of the paper are aware of the limitation of this opposition between Western and non-Western music, which is proposed as an experimental state before the development of further research aimed at the modeling of “stylistic features proper to different stylistic areas”.

As regards the selection of the music collection, some readers might question its composition: its balance between Western and non-Western music, and the way each of these are classified. The selection of genuine non-Western music might be a difficult task as well, perhaps requiring some ethnomusicological expertise. When choosing the non-Western material, the authors discarded musical excerpts “having some

Western influence". In a similar vein, we might wonder whether there could be a need to exclude from the Western collection genres that are deeply influenced by African culture, such as, reggae, blues, and jazz. The selection of representative excerpts within each of these corpora could benefit from an expert's assistance. Moreover, the selection could have been documented in more detail.

The opening of the Western vision of music modeling is founded on a refinement of the Harmonic Pitch Class Profile (HPCP) through an increase of its interval resolution, in order to allow variability in tuning and temperaments. One limitation of this methodology is that octave equivalence is still assumed, causing problems with any music composed of scales with stretched octaves, such as Javanese music (Carterette & Kendall, 1999). Besides, although variability of temperaments are allowed in the modeling of HPCP, some of the features used in the analysis assume equal temperament. These include the tuning frequency, the non-tempered energy ratio (which could be called more precisely non-equally tempered energy ratio), the diatonic strength, the tuning and all the distribution of scale degrees THPCP3, THPCP6, etc. The tuning frequency feature, for instance, is not well-defined if the music is not played in equal temperament. This in turn, is problematic because of the diversity of tunings and scales found in many non-western cultures (Justus & Hutsler, 2005, pp. 12-14). Furthermore, a low value in, say, THPCP3 could mean either that the music is played in equal temperament and the third degree of chromatic scale is infrequently used, or that the music is played in a non-equally tempered scale with no scale degree identical to the third degree of the chromatic scale. An alternative method, not dependent upon an equal tempered scale would consist of inferring and classifying the scale structure directly from the HPCP, with possible expert help from ethnomusicology.

We would like to understand better how the distinction between scale and gamut can be formalized within the HPCP modeling. Several claims concerning the tonal structure of Western music might be questioned by musicologists: the assumption of a rare and unsystematic presence of accidentals in Medieval and Renaissance music might be contradicted by their common use until the sixteenth century of *musica ficta*, i.e., pitch alterations not explicitly indicated in the scores, but nonetheless predictable by expert performers (cf. Huron & Veltman, 2006 for an overview and an empirical assessment of the mode system). Other problematic issues concern the overlooking of minor scales in the characterization of Western music, and the dependence of Western music towards equal temperament and 440 Hz tuning level. Western music that does not follow similar tuning, such as historical performances of baroque music (which is quite in vogue nowadays), might blur the picture and provoke a depreciation of the categorization performance. Indeed, figure 7 shows that a significant part of Western music is not tuned to the 440 Hz diapason, and figure 11 shows that another part has been considered as non-equally tempered.

As explained by the authors, computational approaches to the classification of musical databases into genres and artists have been primarily based on timbral features and rhythmic features. In this latter category, besides tempo and inter-onset-interval ratio, we could also add rhythmic transformation (Guaus & Herrera, 2007) and beat strength (Norowi et al., 2005). Recent studies (Lidy et al., 2007) have discovered ways of potentially improving the results by integrating pitch and tonal features via the automated transcription of recordings into symbolic data. This is where the approach proposed by Gómez and Herrera is of high importance, since it enables an integration of tonal features without the need for error-prone automated transcriptions.

The authors propose a series of features derived from the HPCP that are subsequently used for the discrimination between Western and non-Western music. Some of these features seem to have potential limitations. The non-tempered (or as we said previously, non-equally tempered) energy ratio, for instance, seems to be very sensitive to any slight deviation of the peaks away from the equal-tempered bins in the HPCP, leading to an all-or-nothing measure. Maybe a Gaussian windowing could improve the robustness of the measure? This might also concern the diatonic strength profile, where the actual width of the rectangular windows is not indicated. In order to clarify the concept of octave centroid to the readers, we might add that "octave" should be understood here simply as "register", in terms of absolute position of pitches, as opposed to pitch class dispositions, and has therefore nothing to do with octave equivalence. The roughness estimation, which is actually directly related to the seminal work by Sethares (1998), might be considered as an attempt to generalize the strictly tonal framework in order to integrate features related to timbral aspects of sound. The actual performance of this feature, in particular its distribution along the two databases, would be informative as well. The estimation of degree intensities, given by the features THPCP3, THPCP8 and so on, seems more problematic, as it assumes, as we explained previously, an equal temperament, but also hypothesizes that the main peak in the HPCP can be related to the tonic, and not, for instance, to the dominant. We would be interested to know the general validity of this assumption: in

particular, dominant degrees were called “dominant” as they tend to predominate within the tonal scale. Would it be different in the audio domain due to the sound harmonicity? For all the features, numerous statistical descriptions, such as mean, median, standard deviation, inter-quartile range, kurtosis, and skewness, have been computed, but their use in the classification task has not been detailed.

The statistical description of the feature distribution along the two databases remains somewhat unclear. What is the motivation underlying these computations, and the choice of a variety of statistical distributions? What conclusion do they offer? In particular in Fig. 11, it is difficult to see where the significant difference between the equal-temperament distributions of Western and non-Western music is. Several statistical tests are reported to yield a p-value of zero; to our knowledge this is not theoretically possible with a finite sample of data. A simpler methodology would be to compute the mean and standard deviation for the different features and for the two databases, and then perform a t-test (or its non-parametric variant in the case normality assumption is not met) in order to show the significant differences between Western and non-Western music.

The final section of the article uses the proposed measures in an automated classification task where the pieces are predicted to fall into Western and non-Western categories. One supervised method is based on decision trees, which as a result provides a set of tests enabling a quick (and explicitly presented) classification of the corpora. The resulting tree, displayed in Fig. 12, however displays an absurd configuration, as the latest test performed, concerning the equal temperament of the piece, is actually the one that might need to be performed as early as possible, since most of the other tests were designed for equally tempered music only. The final results of the classification experiments are displayed in Table 2. Musicologists and ethnomusicologists might wonder what conclusion can be drawn from these technical considerations, apart from the fact that computational models, too, are prone to errors. Some examples of misclassified cases could be very informative, as it could help to understand the strengths and weakness of these classifiers. Further research could offer to musicology a more detailed picture concerning in particular the similarity between musical genres in term of tonal content. This could be carried out, for instance, by the use of multiple classes instead of only two for the classification, and the computation of the confusion matrix. We might expect, for instance, that more detailed classification would show interesting relationships between genres that are independent of geographical factors. Furthermore, visualization of the corpora using, say, a Self-Organizing Map (Kohonen, 1997), could provide useful information about the tonal similarities and dissimilarities between different musical cultures.

To support the notion that the Western vs. non-Western distinction is not artificial, the authors carry out unsupervised classification, which, according to them, manages to recombine the Western and non-Western classes from a direct analysis of the audio content, without prior learning. We think, however, that the classification is not completely an “unsupervised learning” process, which would rediscover the categories of Western and non-Western music, since the features used for the classification, as acknowledged by the authors, come from “mathematical models of Western musical scales and consonance”. The selection of features used in the calculation is thus a priori affected by the Western vs. non-Western distinction.

Figure 13 presents the classification results decomposed into musical genres for Western music, and into large geographical areas for non-Western music. We notice that some genres blur the delimitation lines between Western and non-Western: electronica is less Western in terms of tonality, which would not be astonishing after all, and Arabic music is not so non-Western either. The low results for classical music may be due to the probable mixing of very diverse styles, maybe from baroque to contemporary music.

Despite the few shortcomings discussed above, we think the general methodology described in this paper is of high interest. The paper is undoubtedly useful as a brief tutorial on the use of computational methods in comparative music research with large corpora. As such, the paper is rather technical; adding more explanation, examples, and illustrations would probably have increased its accessibility to empirical musicologists. Although we expressed some reservations concerning the feature selection, we are confident that they will be easily overcome in future work. We would like to see more of such strict and systematic testing of hypotheses using empirical methodologies, as such a paradigm could offer valuable insights into the essence of music as a cultural and perceptual phenomenon.

NOTES

[1] Correspondence regarding this study may be sent to Dr. Olivier Lartillot, Department of Music, University of Jyväskylä, Finland, 40014.

REFERENCES

- Carterette, E. C., & Kendall, R. A. (1994). On the tuning and stretched octave of Javanese gamelans. *Leonardo Music Journal*, Vol. 4, 59-68.
- Chordia, P., & Rae, A. (2007, September 23-27). *Automatic Raag classification using pitch-class and pitch-class dyad distributions*. Paper presented at the 8th International Conference on Music Information Retrieval, Vienna, Austria.
- Guaus, E., & Herrera, P. (2007, September 23-27). *A basic system for music genre classification*. Paper presented at the 8th International Conference on Music Information Retrieval, Vienna, Austria.
- Huron, D., & Veltman, J. (2006). A cognitive approach to medieval mode: Evidence for an historical antecedent to the major/minor system. *Empirical Musicology Review*, Vol. 1, 33-55.
- Justus, T., & Hutsler, J. J. (2005). Fundamental issues in the evolutionary psychology of music: Assessing innateness and domain specificity. *Music Perception*, Vol. 23, 1-27.
- Kohonen, T. (1997). *Self-organizing maps*. Berlin: Springer-Verlag.
- Krishnaswamy, A. (2004, October 10-14). *Melodic atoms for transcribing carnatic music*. Paper presented at the 5th International Conference on Music Information Retrieval, Barcelona, Spain.
- Lidy, T., Rauber, A., Pertusa, A., & Iñesta, J. M. (2007). *Improving genre classification by combination of audio and symbolic descriptors using a transcription system*. International Conference on Music Information Retrieval.
- Moelants, D., Cornelis, O., Leman, M., Gansemans, J., De Caluwe, R., De Tré, G., Matthé, T., and Hallez, A. (2006). *Problems and opportunities of applying data and audio mining techniques to ethnic music*. International Conference on Music Information Retrieval.
- Norowi, N. M., Doraisamy, S. C., & Wirza, R. (2005). *Factors affecting automatic genre classification: An investigation incorporating non-western music forms*. International Conference on Music Information Retrieval.
- Sethares, W. A. (1998). *Tuning, Timbre, Spectrum, Scale*. London: Springer-Verlag.