"Solutions in Search of a Problem": Commentary on Videira and Rosa (2017)

PATRICK E. SAVAGE [1]
Keio University Shonan Fujisawa Campus

ABSTRACT: This commentary offers a review of Videira and Rosa's attempt to construct and validate an online corpus of *fado* transcriptions. While I support their application of music information retrieval (MIR) tools to diverse musical repertoires, I fear that a lack of clarity in their goals leads them to fall into the trap of finding "solutions in search of a problem" that is common in computational ethnomusicology. I highlight ways in which I believe they could improve future work on this project, including: 1) more interdisciplinary collaboration, 2) more clarity in their goals, and 3) the use of corpora and tools that are more suitable for comparing the symbolic data contained in their database of musical transcriptions.

IN their article, "A New Online Dataset of Encoded *Fado* Transcriptions", Videira and Rosa present an interesting, albeit imperfect, account of their attempts to construct and validate an online database of musical transcriptions. Although I am a supporter of the authors' goals of applying methods from Music Information Retrieval (MIR) to diverse musical repertoires, I felt that there were still a number of limitations that remain to be addressed.

GOALS

My biggest issue with this article is not a methodological problem, but a frequent lack of clarity about what the authors' motivations and goals are. It is not clear exactly what they are trying to achieve or why they have adopted the methods used. The only explanation appears to a brief reference to "goals of formalising *fado* and building a generative model for its songs." But what exactly does this mean, and why is it important? Without knowing this, it is hard to evaluate whether their method is in fact a useful way of achieving these goals. As such, this looks like potentially yet another example of the kind of "solution in search of a problem" that Tzanetakis, Kapur, Schloss, & Wright (2007) warned against in their review article about the promises and pitfalls of computational ethnomusicology:

> in the majority of existing MIR work that could potentially be used for CE [Computational Ethnomusicology] purposes the authors are primarily engineers or computer scientists, which is not surprising given the early exploratory nature of this area. Unfortunately, frequently this results in a blind application of existing techniques typically to some specific music culture without having a clear musicological goal or motivation. This sometimes results in "solutions in search of a problem". We believe the best way to overcome this is to actively seek interdisciplinary collaborations that include music scholars and technical researchers. Experimental results should generally be interpreted by music scholars with a understanding of the specific music(s) involved, similar to how scientific empiricism and musicological insight can complement each other, as Huron argued (1999). (p. 12)

I am also a firm believer in the value of interdisciplinary collaboration between musicologists and scientists using MIR tools (Savage & Atkinson, 2015; Savage & Brown, 2013), and I also feel that this article and project could benefit from more such collaboration.

MUSICAL SAMPLE

The choice of piano reductions for a project that is supposed to be about vocal songs is curious. I realize that this may have been a strategic choice based on the available sources. However, it is unclear to me which part of the piano reductions is supposed to represent the vocal melody that is presumably of most interest, which part represents the *viola* or *guitarra* accompaniment, and which has been invented for the purpose of the piano reduction. Where would the song lyrics fall? Without knowing these things, it is unclear how useful these transcriptions can be to the goals of studying *fado* songs.

If we accept that the choice of piano reductions is useful, then we need to decide how to sample these reductions. The ad-hoc sampling method described here does not meet usual standards of rigour:

> Since they are sorted arbitrarily, we just encoded all the scores categorized as *fados* along the way until we reached the number of 51, thus rounding up our initial corpus to a total of 100 encoded transcriptions. This number is arbitrary and it was just a convenient sample size given our limited time and budget. There are at least 200 more scores categorized as *fados* in this collection and we intend to encode them as well in the future. (Videira & Rosa, 2017, p. 231)

In such cases, one should usually either sub-sample randomly from the full sample to avoid order effects, or use the entire sample.

ANALYSIS

Rather than using one of the existing MIR tools designed for analysis of symbolic data (i.e., transcriptions) of the kind presented here, the authors made the peculiar choice of converting symbolic data into audio, in some cases even going so far as "using the humaniser preset 'marching band', in the expectation that these MIDI files containing subtle articulatory and gestural variations better reflect a human performance". They then attempt to use MIR tools designed primarily for audio genre classification to perform symbolic genre classification. It appears that after converting their symbolic data into audio they then compare these artificially generated audio files against existing databases of real audio, including in their analysis some features that were absent in the original symbolic data (e.g., dynamics, articulation). This does not seem like a fair comparison. Instead, they should either use MIR tools and databases designed for analysis of symbolic data (e.g., Selfridge-Field, 1995; Urbano, Lloréns, Morato, & Sánchez-Cuadrado, 2011), or they should use a sample of audio recordings of *fado*.

The analysis itself has some issues, although the authors gradually improve their analysis throughout the article. However, much of the analyses they initially report feel like pilot analyses that were clearly flawed and should never really be reported, let alone used to claim misleadingly high levels of classification accuracy (e.g., "huge margins of precision, most of the time, above 90%"). For example, the use of the "marching band" humaniser preset mentioned above is an unjustified extrapolation, and the use of the taxonomy in Table 2 to represent the "world's musics" is unsatisfactory since it contains exclusively Western pop and art music. Their Test 4 is the most well-balanced and controlled test, and this results in a precision of only 49%, so this is really the most reliable value they should report.

The improved taxonomy in Table 3 containing more diverse genres such as "World Beat" and "Western folk" is a good step. I'm particularly interested in the specific comparison they performed between *fado* and flamenco, because these genres seem to me to be stylistically and historically quite similar. It would thus be nice to see some comparison with the extensive MIR work currently being done on flamenco (e.g., Gómez, Díaz-Báñez, Gómez, & Mora, 2014; Guastavino, Gomez, Toussaint, Marandola, & Gomez, 2009). Overall, this represents an interesting and ambitious project, but it requires more refinement and interdisciplinary collaboration to ensure that it finds the best solution for the musical problem(s) of interest.

ACKNOWLEDGEMENTS

NOTES

[1] Correspondence can be addressed to: Patrick E. Savage, Keio University Shonan Fujisawa Campus, Faculty of Environment and Information Studies, 5322 Endo, Fujisawa, Kanagawa 252-0882, Japan, psavage@sfc.keio.ac.jp

REFERENCES

Gómez, F., Díaz-Báñez, J. M., Gómez, E., & Mora, J. (2014). Flamenco music and its computational study. In *Proceedings of Bridges 2014: Mathematics, Music, Art, Architecture, Culture* (pp. 119–126).

Guastavino, C., Gomez, F., Toussaint, G., Marandola, F., & Gomez, E. (2009). Measuring similarity between Flamenco rhythmic patterns. *Journal of New Music Research*, *38*(2), 129–138. https://doi.org/10.1080/09298210903229968

Huron, D. (1999). Music and Mind: Foundations of Cognitive Musicology (The 1999 Ernest Bloch Lectures). Berkeley, CA: University of California. Retrieved from http://www.musiccog.ohio-state.edu/Music220/Bloch.lectures/Bloch.lectures.html

Savage, P. E., & Atkinson, Q. D. (2015). Automatic tune family identification by musical sequence alignment. In *Proceedings of the 16th International Society for Music Information Retrieval Conference (ISMIR 2015)* (pp. 162–168).

Savage, P. E., & Brown, S. (2013). Toward a new comparative musicology. *Analytical Approaches to World Music*, *2*(2), 148–197.

Selfridge-Field, E. (1995). *The Essen musical data package*. Menlo Park, CA: Center for Computer Assisted Research in the Humanities (CCARH).

Tzanetakis, G., Kapur, A., Schloss, W. A., & Wright, M. (2007). Computational ethnomusicology. *Journal of Interdisciplinary Music Studies*, *1*(2), 1–24.

Urbano, J., Lloréns, J., Morato, J., & Sánchez-Cuadrado, S. (2011). Melodic similarity through shape similarity. In *Exploring Music Contents: 7th International Symposium, CMMR 2010* (pp. 338–355). https://doi.org/10.1007/978-3-642-23126-1_21

Videira, T.G., & Rosa, J.M. (2017). A new online archive of encoded fado transcriptions. *Empirical Musicology Review*, *12*(3-4), 229-243. https://doi.org/10.18061/emr.v12i3-4.5431