They compared MFCCs computed from MP3 files at only 32-64 Kbps, observing a decrease in performance when using a different encoder for training and test sets. In contrast, performance did not change significantly when using the same encoder. For genre classification with MFCCs, our results showed no differences in either case. We note though that the bitrates we considered are much larger. Uemura et al. [23] examined the effect of bitrate on chord recognition using chroma features with an SVM classifier. They observed no obvious correlation between encoding and estimation results; the best results were even obtained with very low bitrates for some codecs. Our results on genre classification with chroma largely agree in this case as well; the best results with Lib2 were also obtained by low bitrates. Casey et al. [4] evaluated the effect of lossy encodings on genre classification tasks using audio spectrum projection features. They found a small but statistically significant decrease in accuracy for bitrates of 32 and 96 Kbps. In our experiments, we do not observe these differences, although the lowest bitrate we consider is 64 Kbps. Jacobson et al. [11] also investigated the robustness of onset detection methods to lossy MP3 encoding. They found statistically significant changes in accuracy only at bitrates lower than 32 Kbps.

Our results showed that MFCCs and chroma features, as computed by Lib1 and Lib2, are generally robust and stable within reasonable limits. Some differences have been noted between tools though, largely attributable to the different frequency ranges they employ. Nonetheless, it is evident that certain combinations of codec and bitrate may require a re-parameterization of some descriptors to improve or even maintain robustness. In practice, these parameterizations affect the performance and applicability of algorithms, so a balance between performance, robustness and generalizability should be sought. These considerations are of major importance when collecting audio files for some dataset, as a minimum audio quality might be needed for some descriptors.

7. CONCLUSIONS

In this paper we have studied the robustness of two common audio descriptors used in Music Information Retrieval, namely MFCCs and chroma, to different audio encodings and analysis parameters. Using a varied corpora of music pieces and two different audio analysis tools we have confirmed that MFFCs are robust to frame/hop sizes and lossy encoding provided that a minimum bitrate of approximately 160 Kbps is used. Chroma features were shown to be even more robust, as the codec and bitrates had virtually no effect on the computed descriptors. This is somewhat expected given that chroma does not capture information as fine-grained as MFCCs do, and that lossy compression does not alter the perceived tonality. We did find subtle differences between implementations of these audio features, which call for further research on standardizing algorithms and parameterizations to maximize their robustness while maintaining their effectiveness in the various tasks they are used in. The immediate line for future work includes the analysis of other features and tools.

8. ACKNOWLEDGMENTS

This work is partially supported by an A4U postdoctoral grant and projects SIGMUS (TIN2012-36650), CompMusic (ERC 267583), PHENICX (ICT-2011.8.2) and GiantSteps (ICT-2013-10).

9. REFERENCES

[1] J.J. Aucouturier, F. Pachet, and M. Sandler. "The way it sounds": timbre models for analysis and retrieval of music signals. *IEEE Trans. Multimedia*, 2005.

[2] D. Bogdanov, N. Wack, et al. ESSENTIA: an audio analysis library for music information retrieval. In *ISMIR*, 2013.

[3] C. Cannam, M.O. Jewell, C. Rhodes, M. Sandler, and M. d'Inverno. Linked data and you: bringing music research software into the semantic web. *J. New Music Res.*, 2010.

[4] M. Casey, B. Fields, et al. The effects of lossy audio encoding on genre classification tasks. In *AES*, 2008.

[5] W. Chai. Semantic segmentation and summarization of music: methods based on tonality and recurrent structure. *IEEE Signal Processing Magazine*, 2006.

[6] D. Ellis. Classifying music audio with timbral and chroma features. In *ISMIR*, 2007.

[7] T. Fujishima. Realtime chord recognition of musical sound: a system using common lisp music. In *ICMC*, 1999.

[8] T. Ganchev, N. Fakotakis, and G. Kokkinakis. Comparative evaluation of various MFCC implementations on the speaker verification task. In *SPECOM*, 2005.

[9] E. Gómez. *Tonal description of music audio signals*. PhD thesis, Universitat Pompeu Fabra, 2006.

[10] S. Hamawaki, S. Funasawa, et al. Feature analysis and normalization approach for robust content-based music retrieval to encoded audio with different bit rates. In *MMM*, 2008.

[11] K. Jacobson, M. Davies, and M. Sandler. The effects of lossy audio encoding on onset detection tasks. In *AES*, 2008.

[12] J.H. Jensen, M.G. Christensen, D. Ellis, and S.H. Jensen. Quantitative analysis of a common audio similarity measure. *IEEE TASLP*, 2009.

[13] B. McFee, L. Barrington, and G. Lanckriet. Learning content similarity for music recommendation. *IEEE TASLP*, 2012.

[14] D.C. Montgomery. *Design and Analysis of Experiments*. Wiley & Sons, 2009.

[15] M. Müller and S. Ewert. Towards timbre-invariant audio features for harmony-based music. *IEEE TASLP*, 2010.

[16] M. Müller, H. Mattes, and F. Kurth. An efficient multiscale approach to audio synchronization. In *ISMIR*, 2006.

[17] J. Paulus, M. Müller, and A Klapuri. Audio-based music structure analysis. In *ISMIR*, 2010.

[18] L.R. Rabiner and R.W. Schafer. *Introduction to Digital Speech Processing*. Foundations and Trends in Signal Processing. 2007.

[19] J. Reed and C. Lee. Preference music ratings prediction using tokenization and minimum classification error training. *IEEE TASLP*, 2011.

[20] J. Serrà, E. Gómez, and P. Herrera. Audio cover song identification and similarity: background, approaches, evaluation, and beyond. In Z. Raś and A.A. Wieczorkowska, editors, *Advances in Music Information Retrieval*. Springer, 2010.

[21] S. Sigurdsson, K.B. Petersen, and T. Lehn-Schiler. Mel Frequency Cepstral Coefficients: an evaluation of robustness of MP3 encoded music. In *ISMIR*, 2006.

[22] M. Slaney. Auditory toolbox. *Interval Research Corporation, Technical Report*, 1998. `http://engineering.purdue.edu/~malcolm/interval/1998-010/`.

[23] A. Uemura, K. Ishikura, and J. Katto. Effects of audio compression on chord recognition. In *MMM*, 2014.

[24] K. Yoshii, M. Goto, K. Komatani, T. Ogata, and HG. Okuno. An efficient hybrid music recommender system using an incrementally trainable probabilistic generative model. *IEEE TASLP*, 2008.