Impact of audio codec and quality on genre classification and BPM recognition in Essentia Sjoerd Hulleman s.hulleman@student.tudelft.nl Supervisors: Cynthia C.S. Liem, Jaehun Kim



Background

- Do machine learning procedures which extract audio features actually extract musically meaningful information?
- Audio quality is not often considered and reported in research setups
- Could have an impact on audio feature extraction.
- Essentia is an open-source C++ library for audio analysis and audio-based music information re-trieval.
- Impact on rock genre highest.
- BPM not explicitly researched in earlier work.

Research questions

How do different audio codecs and audio quality impact genre classification and beats per minute (BPM) recognition in Essentia?

- What is the performance of genre classification with FLAC encoding?
- What is the performance of BPM recognition with FLAC encoding?
- How is genre classification influenced by MP3 quality/bitrates?
- How is BPM recognition influenced by MP3 quality/bitrates?
- How is genre classification influenced by WAV quality/bitrates?
- How is BPM recognition influenced by MP3 quality/bitrates?

Method

To obtain music files in their original format, we collaborated with Muziekweb. They are run by music professionals, among which musicologists. All their albums are labeled with their genre, this is used to hand-pick 66 albums as a dataset for this research.



Results

 $\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$

Mean squared error used as measure, tables visualized in the $\frac{n}{2}$ boxplots on the right, using the log (base 10) squared error.

Audio codec	MSE rock probability	Audio codec	MSE BPM
FLAC	0	FLAC	0
MP3 320 kbps	0.000001	MP3 320 kbps	8 18.640363
MP3 256 kbps	0.000002	MP3 256 kbps	5 27.831408
OGG 320 kbps	0.000024	OGG 320 kbp	s 29.732891
OGG 128 kbps	0.000054	MP3 128 kbps	37.494367
MP3 128 kbps	0.000449	OGG 128 kbp	s 66.388747
MP3 96 kbps	0.001160	MP3 96 kbps	68.39757
OGG 96 kbps	0.001369	OGG 96 kbps	75.439119
OGG 64 kbps	0.001511	2 GG 64 kbps	112.913352



$log_{10}((Y_i-\hat{Y}_i)^2)$

- FLAC used as the baseline for MSE and squared error.
- MSE keeps going up when the audio quality further degrades. This is the case for rock classification as well as BPM estimation.
- Clear variation between codecs for rock classification as well as BPM estimation.
- The order of codecs is eqactly the same for the results of rock classification and BPM estimation, while these are separate algorithms.
- BPM has an "explosive" growth in MSE, this is due to BPM difference being relatively large compared to rock probability difference (e.g. 0.15 compared to 12).

Conclusions

- Audio codecs and quality have a clear influence on the results in MIR.
- Audio codecs and quality should be more explicitly considered in research setups
- No conclusion on which audio codec is best, more data is needed for this.