OPEN BROADCAST MEDIA AUDIO FROM TV: A DATASET OF TV BROADCAST AUDIO WITH RELATIVE MUSIC LOUDNESS ANNOTATIONS

Blai Meléndez-Catalán[1,2], Emilio Molina[2], Emilia Gómez[1,3]
[1]MTG, Universitat Pompeu Fabra, Barcelona (Spain)
[2]BMAT Licensing S.L., Barcelona (Spain)
[3]Joint Research Centre, European Comission
{bmelendez,emolina}@bmat.com, emilia.gomez@upf.edu

EXTENDED ABSTRACT

Music detection refers to the task of finding music segments in an audio file.[1] The annotations about the presence of music are the minimum requirement for a dataset to be suitable for this task. However, the following two features are essential to any music detection dataset that aims to provide a certain level of generalization: first, music should appear both isolated and mixed with other type of non-music sound., Otherwise, the dataset may not be representative of real-life scenarios such as broadcast audio; and second, a significant number of the audio files should be multi-class, i.e., contain class changes to allow the evaluation of an algorithm's precision in detecting them.

The main application of music detection algorithms is the monitoring of music for copyright management. In the current business model, broadcasters are taxed based on the percentage of music they broadcast. Additionally, it is important to know whether this music is used in the foreground or the background as it is considered differently for the distribution of copyright royalties by some collective management organizations.[2] In this scenario, we need to estimate the loudness of music in relation to other simultaneous non-music sounds, i.e., its relative loudness. We define *relative music loudness estimation* as the task of finding music segments in an audio file and classifying them into foreground or background music.

Open Broadcast Media Audio from TV (OpenBMAT) is the only open dataset for the task of music detection that brings together all the appropriate characteristics for this task and also for the task of estimating the music's relative loudness. OpenBMAT contains 27.4 hours of audio divided in 1647 one-minute audio files. Each of these audio files comes from a different recording that we have sampled from BMAT's[3] private database, which temporarily stores recordings from over 2000 TV channels that this company monitors as part of its business. We consider that having many short audio files allows the dataset to include a greater variety of contexts. Nevertheless, these audio files are long enough to be multi-class.

The taxonomy that we use to annotate these audio files is formed by 6 classes. Each of these classes applies to a different combination, in terms of loudness, of content that is considered music and content that is not. The classes are: *Music*, *Foreground Music*, *Similar*, *Background Music*, *Low Background Music* and *No Music*.

OpenBMAT has been manually cross-annotated by three different annotators. Cross-annotating allows us to assess the reliability of the annotations produced. The tool that we have used for the annotation of the dataset is BAT[4] [2], an open-source, web-based tool for the manual annotation of events in audio files.

[1] https://www.music-ir.org/mirex/wiki/2018:Music_and/or_Speech_Detection
[2] https://createurs-editeurs.sacem.fr/brochures-documents/regles-de-repartition-2017
[3] https://www.bmat.com/
[4] https://github.com/BlaiMelendezCatalan/BAT

To assess the reliability of the annotations we use the percentage of agreement between the three annotators in the annotated classes. We define two different levels of agreement: full agreement, which happens when all three annotators have annotated the same class; and partial agreement, which happens when at least two annotators have annotated the same class. These values can be computed considering all the classes in the taxonomy, but also for two mappings of these classes. These mappings adapt the original taxonomy to the tasks of music detection (MD) and relative music loudness estimation (RMLE). The MD mapping unifies all classes under the *Music* class except for the *No Music* class, which is left unchanged. In the *R*MLE mapping, the *No Music* class also remains unchanged, the *Music* and *Foreground Music* classes are merged into the *Foreground Music* class and the remaining classes are mapped to the *Background Music* class.

When considering all classes, there is already a percentage of partial agreement of 96.75%. This percentage increases to 99.79% with the RMLE mapping. The percentage of full agreement when considering all classes is 68.18%, which increases to 89.1% for the RMLE mapping and to 94.78% for the MD mapping. Over 35% of the audio files have a percentage of full agreement higher than 99% and for almost 90% of the audio files it is higher than 70%.

The content distribution in terms of classes for each annotator for the complete taxonomy and both mappings is similar: first, all annotators have considered that around 50% of the dataset is *No Music*, which implies that the other 50% contains music that appears either isolated or mixed with non-music sounds. This means that OpenBMAT is balanced in terms of music and non-music content. Second, the part of the dataset containing music is approximately distributed with a 30%/70% proportion between the RMLE mapping *Foreground Music* and *Background Music* classes, respectively.

We have evaluated a state-of-the-art music detection algorithm with OpenBMAT with two goals in mind: first, to find the most challenging content in OpenBMAT; and second, to highlight the importance of the agreement information. The selected algorithm is the winner of 2018 MIREX[5] competition for the task of music detection [1]. Related to the first goal, there are three groups of estimation errors: (1) those related to the ambiguity of what is music and what is not; (2) those related to the heterogeneity of music and non-music sounds; and (3) those related to the music's volume. Regarding the second goal, agreement information indicates what segments are too ambiguous for three humans to agree, and might not be suitable to train an algorithm or evaluate its performance. Also, it allows us to distinguish between the errors of an algorithm that happen in audio files with a clear ground truth and in audio files with an ambiguous ground truth. Potential users will probably find more beneficial to focus in the first type of errors in order to improve their algorithms.

This extended abstract is a summary of a paper [3] published in the TISMIR Journal[6].

ACKNOWLEDGMENTS

We thank the annotators for the effort they invested in the annotation of OpenBMAT. We also thank Alex Ciurana from BMAT with which we have discussed about issues essential to the construction of this dataset. We thank Brisa Burriel for her advice on the best way to make OpenBMAT available to the research community and Sonia Espi to provide the means to do so. Finally, we thank the Catalan Industrial Doctorates Plan for funding this research.

REFERENCES

[1] Blai Meléndez-Catalán. Music and/or Speech Detection MIREX 2018 Submission, 2018. Music Information Retrieval Evaluation eX-change.

[2] Blai Meléndez-Catalán, Emilio Molina, and Emilia Gómez. Bat: An open-source, web-based audio events annotation tool. In *3rd Web Audio Conference*, 2017.

[3] Blai Meléndez-Catalán, Emilio Molina, and Emilia Gómez. Open broadcast media audio from tv: A dataset of tv broadcast audio with relative music loudness annotations. *Transactions of the International Society for Music Information Retrieval*, 2(1):43–51, 2019.

[5] https://www.music-ir.org/mirex/wiki/MIREX_HOME
[6] https://transactions.ismir.net/