Co-creative Musical Repurposing by Modelling Rhythmic Compatibility

André C. Santos
CISUC, DEI
University of Coimbra, Portugal
andresantos@dei.uc.pt

Matthew E. P. Davies
CISUC, DEI
University of Coimbra, Portugal
mepdavies@dei.uc.pt

Amílcar Cardoso
CISUC, DEI
University of Coimbra, Portugal
amilcar@dei.uc.pt

Abstract

A common, and arguably innate, human response when listening to music is to tap one's foot to mark the regular pulse of the beat. A more complex form of interactive synchronization occurs when listeners tap out rhythmic patterns using their fingers, hands, or even some form of improvised drumsticks. The proposed goals of this research are two-fold: i) to investigate this complex, but under-explored, phenomenon of rhythmic engagement by building a computational model of rhythmic compatibility; and ii) to leverage this knowledge to drive novel means for co-creative content repurposing via the layering and integration of user-tapped rhythms in a timbrally-consistent way with the source audio. While this research is predominantly computational, it will be approached in a multi-disciplinary fashion drawing upon music theory, psychology, computational creativity, and machine learning in pursuit of musically meaningful next-generation tools for enhanced creativity and content personalization.

Related Work

This research has its roots in exploratory work on "interactive redrumming" (López-Serrano et al. 2018), and its successful attainment depends on the novel extension and combination of several research topics within the field of music information retrieval (MIR), namely: Automatic Drum Transcription (ADT) (Vogl 2018; Wu et al. 2018) - which seeks to identify the temporal location of percussion events and label them according to a fixed drum vocabulary; Musical Source Separation (MSS) (Défossez 2021), - which targets the separation of mixed recordings into isolated channels corresponding to each instrument; and music compatibility - which tries to model how musical parts are related to one another and can be combined in a musically coherent manner (Chen, Smith, and Yang 2020).

While all three are independent research topics, they have each been positively impacted and advanced by the widespread uptake of deep learning within MIR. Both ADT and MSS have benefitted from the use of supervised learning via deep convolutional and recurrent neural networks (Vogl 2018; Défossez 2021) which have superseded unsupervised approaches built on non-negative matrix factorization (Wu et al. 2018). Be that as it may, persistent challenges remain in generalization to unseen and/or under-represented instrument types (Cartwright and Bello 2018), which have led to novel deep learning approaches including data augmentation (Rohit M.A., Bhattacharjee, and Rao 2021; Cartwright and Bello 2018; Vogl 2018), few-shot learning (Wang et al. 2020), and content-specific adaptation of pre-trained models (Davies et al. 2020; Rohit M.A., Bhattacharjee, and Rao 2021).

Within this research, these issues are highly pertinent given the need to transcribe user-tapped rhythms captured in unconstrained recording conditions and with arbitrary "drum" timbres. In the absence of explicitly labelled targets for user-tapped rhythms, the goal for transcription instead becomes the musically meaningful association with drum timbres present in the source material, and thus targets a model of rhythmic compatibility. The association and transformation of tapped-to-performed rhythms in a multi-timbral context has never been addressed and thus has the potential to significantly advance the current state of the art (Chen, Smith, and Yang 2020).

Objectives

The principal objectives of this research are as follows:

1. To develop an automatic percussion transcription system for tapped rhythms with arbitrary timbre and unconstrained recording conditions. Realizing this objective will enable the transcription of tapped rhythms acquired without specialized hardware, i.e., using laptop microphones. A particular challenge will be to contend with the wide range of non-standard timbres for percussive events.

2. To build a multi-timbral model of rhythmic compatibility. This objective first targets the precise temporal alignment of the tapped rhythms with those present in the source recordings, and then seeks to identify commonalities in the rhythmic structures based on the distribution of timbrally-distinct events.

3. To develop a novel method for timbre-transformation of tapped rhythms conditioned on the musical audio source. Leveraging state-of-the-art techniques for musical audio source separation to isolate the percussion components of the source material, this objective targets the timbre-transformation of the tapped rhythms to render them such

that they are timbrally consistent with the source. Moreover, with rhythmic compatibility assured, we can extract different timbres from different sources and use this loop library for creative activities like remixing or loop-based music.

4. To investigate user engagement with musical content repurposing driven by rhythmic interaction as a co-creative effort. This objective targets the user-based evaluation of the computational aspects of this project towards the measurement and understanding of the level of musical engagement by tapping rhythms and experiencing the results mixed back in with the source material.

Detailed Description

Automatic Transcription of Tapped Rhythms: The first research area concerns the acquisition and transcription of tapped rhythms which are made while a user is actively listening to music. Unlike existing work in ADT, which typically addresses professional recordings and expects a fixed drum vocabulary, we operate from a fundamentally different perspective. First, signal acquisition will be made without high-quality equipment and in unconstrained conditions, e.g., with a laptop microphone while the user listens to musical content on headphones to prevent leakage. Second, the "drum" timbres employed by the user will be considered completely arbitrary and could result from a variety of surfaces and different means of striking them. In this way, the initial transcription will focus on precise timing and categorization into distinct timbres via deep unsupervised learning (Choi and Cho 2019). In the event this proves unreliable, we will explore approximate timbral calibration as in (Ramires 2017) towards a "human-in-the-loop" transcription.

Rhythmic Compatibility Model: A critical and under-explored challenge in the mixing and repurposing of music content relates to rhythm. In existing work this is largely restricted to the alignment of the beat and downbeat structure ("beat matching"), with little research considering compatibility among rhythmic patterns. The paradigm of rhythmic engagement via tapped rhythms provides a direct means to observe compatible rhythms and musically meaningful variations thereof. Relating the metrical structure from the source material to the tapped rhythms, e.g., using the current state of the art (Böck and Davies 2020), and leveraging ADT of the source, will provide the basis for a computational model of rhythmic compatibility. This model will explore and map correspondences between the arbitrary timbres in the tapped rhythms with those in the source material based, as well as leveraging similarity in so-called "reduced" (i.e., predominant) rhythmic patterns derived from a generative model of rhythm characterization (Sioros, Davies, and Guedes 2018).

Percussive Timbre Transfer: Building upon the rhythmic compatibility model, this task investigates rendering the user tapped patterns with the drum timbres of the musical source material. Extending recent research in rhythm transformation and drum synthesis (Tomczak, Goto, and Hockman 2020) its goal is to provide a high-quality resynthesis of the tapped rhythm to be mixed back with the source material. Obtaining high quality target drum timbres from the source material will depend on the use of state-of-the-art MSS techniques parameterized to minimize interference for drum parts, potentially at the expense of the separation quality of other sources. To allow user customization of the rendered timbres, our approach will allow exploration in the latent space of the generative networks used for rhythm transformation.

Co-Creative Repurposing of Musical Content: The final task in this research directly considers the question of user engagement via experimentation with co-creative music technology. It will be centered on the design, execution, and analysis of a user study in which participants experiment with tapping rhythms to music content and experiencing their results rendered and layered with the source material. Of particular importance will be understanding the extent to which participants report a sense of being creative in this task, and whether this activity aided in users' ability to both devise and "perform" new rhythms which they might otherwise be unable to conceive. (Liapis et al. 2016)

Current Work

Since the beginning of the PhD, a dataset of 17 songs was collected, and tapped rhythms were recorded for each one of those songs. Different genres and drum styles were selected towards building a robust system, but song familiarity was also taken into consideration to have a reliable tapped version of each song's drums/percussion.

At ISMIR 21 last October, Rohit et al. (Rohit M.A., Bhattacharjee, and Rao 2021) presented an adapted ADT tool (Vogl 2018) to transcribe Tabla solos. By exploiting the advantages of data augmentation, they built a deep supervised learning model with less than two hours of tabla solo recordings. This seems to be a very promising result, so we decided to try and adapt this approach to our specific case.

To this end, we need to build our labeled dataset by performing ADT on our original songs dataset. We also leveraged MSS tools to remove the songs' original drum tracks and, using one shot drum hits to synthesize drum sounds from the drum transcriptions, we reconstructed our song dataset to evaluate the transcription quality.

As of now, the transcribed dataset needs to be "cleaned up", i.e, we need to manually edit and rectify any transcription errors so that these transcriptions can then be used to train a deep supervised model to perform tapped rhythm transcription. We also intend to explore data augmentation methods and different model architectures in the training stage. Finally, we need to test the model to assess the overall quality of the transcriptions.

As a contingency measure, we will explore the use of deep unsupervised approaches (Choi and Cho 2019). We also believe it to be useful to transcribe the tapped rhythms into some symbolic representation, but this extra step might introduce some unnecessary error for our ultimate goal which is to render tapped rhythm information as high quality drum sounds, so if this first approach proves untrustworthy, we intend to explore an audio-to-audio "translation", directly converting the arbitrary timbered tapped rhythms into drum sounds.

Acknowledgements

This work is funded by national funds through the FCT - Foundation for Science and Technology, I.P., within the scope of the project CISUC - UID/CEC/00326/2020 and by European Social Fund, through the Regional Operational Program Centro 2020. The author is funded by Foundation for Science and Technology (FCT), Portugal, under the grant 2021.05653.BD.

References

Böck, S., and Davies, M. E. P. 2020. Deconstruct, analyse, reconstruct: How to improve tempo, beat, and downbeat estimation. In *Proceedings of the 21st International Society for Music Information Retrieval Conference*, 12–16. ISMIR.

Cartwright, M., and Bello, J. P. 2018. Increasing drum transcription vocabulary using data synthesis. In *Proceedings of the International Conference on Digital Audio Effects*, 72–79. DAFx.

Chen, B.-Y.; Smith, J. B.; and Yang, Y.-H. 2020. Neural loop combiner: Neural network models for assessing the compatibility of loops. In *Proceedings of the 21st International Society for Music Information Retrieval Conference*, 424–431. ISMIR.

Choi, K., and Cho, K. 2019. Deep unsupervised drum transcription. In *Proceedings of the 20th International Society for Music Information Retrieval Conference*, 183–191. ISMIR.

Davies, M. E. P.; Fuentes, M.; Fonseca, J.; Aly, L.; Baraldi, F. B.; and Jerónimo, M. 2020. Moving in time: Computational analysis of microtiming in maracatu de baque solto. In *Proceedings of the 21st International Society for Music Information Retrieval Conference*, 795–802. ISMIR.

Défossez, A. 2021. Hybrid spectrogram and waveform source separation. In *Proceedings of the ISMIR 2021 Workshop on Music Source Separation*. ISMIR.

Liapis, A.; Yannakakis, G. N.; Alexopoulos, C.; and Lopes, P. L. 2016. Can computers foster human users' creativity? theory and praxis of mixed-initiative co-creativity. *Digital Culture & Education* 8.

López-Serrano, P.; Davies, M. E.; Hockman, J.; Dittmar, C.; and Müller, M. 2018. Break-informed audio decomposition for interactive redrumming. In *Late Breaking and Demo Session of the 19th International Society for Music Information Retrieval Conference*. ISMIR.

Ramires, A. F. S. 2017. Automatic transcription of drums and vocalised percussion. Master's thesis, FEUP - Faculdade de Engenharia da Universidade do Porto.

Rohit M.A.; Bhattacharjee, A.; and Rao, P. 2021. Four-way Classification of Tabla Strokes with Models Adapted from Automatic Drum Transcription. In *Proceedings of the 22nd International Society for Music Information Retrieval Conference*, 19–26. ISMIR.

Sioros, G.; Davies, M. E. P.; and Guedes, C. 2018. A generative model for the characterization of musical rhythms. *Journal of New Music Research* 47(2):114–128.

Tomczak, M.; Goto, M.; and Hockman, J. 2020. Drum synthesis and rhythmic transformation with adversarial autoencoders. In *Proceedings of the 28th ACM International Conference on Multimedia*, 2427–2435. ACM-MM.

Vogl, R. 2018. *Deep learning methods for drum transcription and drum pattern generation*. Ph.D. Dissertation, Johannes Kepler University Linz, Linz.

Wang, Y.; Salamon, J.; Cartwright, M.; Bryan, N. J.; and Bello, J. P. 2020. Few-shot drum transcription in polyphonic music. In *Proceedings of the 21st International Society for Music Information Retrieval Conference*, 117–124. ISMIR.

Wu, C.-W.; Dittmar, C.; Southall, C.; Vogl, R.; Widmer, G.; Hockman, J.; Müller, M.; and Lerch, A. 2018. A review of automatic drum transcription. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 26(9):1457–1483.