Adapting Beat Tracking Models for Salsa Music: Establishing a Baseline with a novel dataset

Antonin Rapini[1*] and Anna Jordanous[2]

[1*]School of Computing, University of Kent, United Kingdom, aplr3@kent.ac.uk
[2]School of Computing, University of Kent, United Kingdom

Abstract— **This study addresses the challenge of adapting current beat tracking algorithms, predominantly trained on Western music, to the rhythmic complexities of Salsa, a genre rich in syncopations and polyrhythms. We benchmark the adaptability of three established models: BeatNet, Wavebeat, and Böck TCN, using our own newly introduced beat-annotated Salsa dataset and focusing on training methods that minimize the need for extensive annotated data. We find that, on Salsa music, models trained with popular datasets and fine-tuned with Salsa generally outperform models trained under other training conditions. This research not only establishes a baseline for beat tracking performance in Salsa music but also contributes to the broader goal of developing more universally adept music information retrieval systems.**

I. BACKGROUND

Beat tracking is the temporal identification of "beats", the basic rhythmic unit of a song. Although it is a skill that comes naturally for many people, automatic beat tracking, the programmatic identification of beats from audio data, is a highly complex task. Current state of the art beat tracking algorithms perform well on Western music [2] but often stumble when encountering the rhythmic intricacies of Salsa. Salsa is, a genre rich with syncopation and polyrhythms [4], features not often found in current available beat tracking datasets. Addressing this discrepancy is crucial for progress towards more universal music information retrieval.

II. OBJECTIVES

This study sets out to benchmark the adaptability of beat tracking models to Salsa music. It makes a point to emphasise methods that circumvent the need for large amounts of annotated data, a rare commodity that requires hours of tedious work and expert knowledge to be created.

III. METHODOLOGY

We assess the accuracy of three prominent beat tracking models: BeatNet [1], Wavebeat [2], and Böck TCN [3], on an unseen Salsa test dataset created for this study. The dataset contains 40 songs for a total of 2h53 of beat-annotated music. It will be made available at github.com/AntoninRap/Salsa-dataset. The models were trained under four distinct conditions: training on "other"* datasets of non-Salsa music, training exclusively on Salsa, an initial training on "other" datasets followed by fine-tuning on Salsa, and simultaneous training on both Salsa and "other" datasets.

*"Other" refers to four datasets of mostly western music popular in the field of beat tracking: GTZAN, Rock, Ballroom and SMC

IV. RESULTS

Our findings reveal a consistent pattern: models fine-tuned with Salsa music outperformed those trained on more generalised datasets. This trend persisted across models: Models trained on "other" datasets then fine-tuned on Salsa achieved the highest accuracy, followed by those trained on a combination of Salsa and "other" datasets, then models trained only on Salsa, with the least accuracy seen in models trained exclusively on "other" datasets. These results are still generally below what we see on "other" genres.

V. CONCLUSION

The study demonstrates that fine-tuning beat tracking models with genre-specific data can significantly improve accuracy for Salsa music. It also establishes a baseline for the performance of beat tracking on this genre, providing a reference point for the efficacy of more intricate future methodologies. This work contributes to the ongoing efforts to develop beat tracking systems that better account for the rhythmic diversity found in global music genres, and for that goal, introduces a new beat-annotated dataset of Salsa music.

Table 1. average f-measure accuracy on unseen Salsa test dataset (10 songs) of three prominent beat tracking models under the four training conditions outlined in the Methodology section.

F-measure accuracy				
Model	***Fine-tuned***	***Salsa + others***	***Salsa only***	***Others***
BeatNet	**0.844**	0.807	0.710	0.539
TCN	0.789	0.644	0.461	0.422
Wavebeat	0.739*	*	0.789	0.704*

* Initial results with base model ("other") not trained by us, final results still pending

REFERENCES

[1] Heydari, Mojtaba, Frank Cwitkowitz, and Zhiyao Duan. "BeatNet: A real-time music integrated beat and downbeat tracker." International Society for Music Information Retrieval (2021).

[2] Steinmetz, Christian J., and Joshua D. Reiss. "WaveBeat: End-to-end beat and downbeat tracking in the time domain." arXiv preprint arXiv:2110.01436 (2021).

[3] Matthew Davies, E. P., and Sebastian Böck. "Temporal convolutional networks for musical audio beat tracking." 2019 27th European Signal Processing Conference (EUSIPCO). IEEE, 2019.

[4] Laura M. Getz, Scott Barton & Lynn K. Perry (2021) Context-specific Knowledge Is the "Key" to Salsa Music, Auditory Perception & Cognition, 4:1-2, 14-32, DOI: 10.1080/25742442.2021.1964341