DUAL-ONSET DETECTION FOR RHYTHMIC ANALYSIS OF JAZZ RECORDINGS

Dave Foster[†] *and Simon Dixon*

C4DM, Queen Mary University of London
d.foster@qmul.ac.uk

ABSTRACT

Within the research topic of ensemble jazz timing, the theories of Participatory Discrepancies (PD) [1] loom large. The micro-rhythmic interplay between musicians is a large factor in engendering the "swing" feel: which is as elusive when trying to describe qualitatively as it is to teach. The PD theories, particularly in relation to jazz accompaniment [2], have been tested in small scale studies, using limited numbers of hand-annotated examples [3], but never in a large corpus study, due to the lack of sufficiently annotated data in which the individual onsets of each instrument are disentangled from the composite mix.

Recent developments in source separation [4] and onset detection [5] tease the possibility of automatically annotated datasets, where instrumental tracks are de-mixed from the blend and onsets derived from the resultant separated audio. But how accurate and reliable are these methods and architectures for this specific task? Are the loss functions employed for training and the metrics used for evaluation suitable for this specific task?

Using a dataset of jazz quartet recordings, where stems of each instrument are available, we first test the effectiveness of existing systems by comparing the onsets derived from the original tracks to those derived from audio separated from a mixture. The source separation architectures are largely trained using pop/rock music (hence with bass guitar and drums), so we then attempt to make improvements by re-training on the target data (double bass and drums).

We propose and test some novel approaches and pipelines, including an end-to-end focused dual-onset detector for identifying bass and drum onsets in the region of a known beat event. The bass/drums timing interaction within the training dataset is analysed, with reference to the PD proposals, albeit with the constraint that the music was recorded with a fixed tempo click track (and hence no micro- or macro- tempo deviations are possible).

We then apply the best performing dual-onset architectures to the accompaniment tracks of the Filosax dataset [6] (originally from Jamey Aebersold[1] backing tracks), allowing for the first big data investigation of the PD proposals. These tracks are not recorded with a click track, so the fluctuations in tempo can be viewed from the perspective of any member of the ensemble, or via the interplay between them.

The resultant values also give an additional level of detail to the timing annotations of the Filosax data itself, where previously beat times were derived from the amalgamated rhythm section tracks. As a result, we are able to suggest ways in which these accurate accompaniment timing values can be used to model and predict the performance timings of an interacting soloist.

[†]UKRI Centre for Doctoral Training in Artificial Intelligence and Music, supported by UK Research and Innovation [grant number EP/S022694/1].

[1] http://jazzbooks.com/jazz/JBIO

1. REFERENCES

[1] C. Keil, "Participatory discrepancies and the power of music," *Cultural Anthropology*, vol. 2, no. 3, pp. 275–283, 1987.

[2] J. A. Prögler, "Searching for swing: Participatory discrepancies in the jazz rhythm section," *Ethnomusicology*, vol. 39, no. 1, pp. 21–54, 1995.

[3] M. Butterfield, "Participatory discrepancies and the perception of beats in jazz," *Music perception*, vol. 27, no. 3, pp. 157–176, 2010.

[4] S. Rouard, F. Massa, and A. Défossez, "Hybrid transformers for music source separation," in *International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2023, pp. 1–5.

[5] M. Mounir, P. Karsmakers, and T. van Waterschoot, "Musical note onset detection based on a spectral sparsity measure," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2021, pp. 1–17, 2021.

[6] D. Foster and S. Dixon, "Filosax: A dataset of annotated jazz saxophone recordings," in *22nd International Society for Music Information Retrieval Conference*, 2021.