Adapting Meter Tracking Models to Latin American Music

Lucas S. Maia[1], Martín Rocamora[2], Luiz W. P. Biscainho[1], Magdalena Fuentes[3]

[1] PEE/COPPE, Federal University of Rio de Janeiro, Brazil
[2] FING, Universidade de la República, Uruguay
[3] MARL-IDM, New York University, United States

lucas.maia@smt.ufrj.br

Introduction

Goal: adapt state-of-the-art beat and downbeat tracking models based on temporal convolutional networks [1] to Latin American Music with small quantities of data

Target: two Latin American music traditions—Uruguayan *candombe* and Brazilian *samba*

Hypothesis: TCN model achieves good performance in *candombe* and in *samba* with little data due to high genre homogeneity

Proposed analysis:

- train models with increasing amounts of annotated data (< 1 min up to ~40 min)
- use three training strategies (FS, FT, and data augmentation)
- contrast model performance against off-the-shelf models trained in/developed for Western music
- contrast model performance against an effective baseline, a Bayesian model (BayesBeat) [2]

Datasets

Candombe dataset [3]:

- 35 rec. of *candombe* drumming (2.5 h)
- 4/4 meter with syncopation
- *chico*, *repique*, and *piano* drums
- irregular rhythm cycle division

Samba dataset (BRID) [4]:

- 93 short tracks, 30 s each (0.74 h)
- 2/4 meter
- ten different instrument classes: *tamborim*, *pandeiro*, *surdo*, *cuíca*, *agogô*, among others
- complex combination of timbres
- strong accent at second beat

Small Data

- Segment *Candombe* into 30-second snippets
- For both genres: select 93 excerpts, split in train / test (80% / 20%)
- **Train**: 10 s at the beginning of each excerpt (5 s train and 5 s validation)
- **Test**: full excerpts
- Use training subsets of increasing size:
  - 4, 9, 18, 37, 55, and 74 excerpts
  - Equivalently: 0.67, 1.5, 3, 6, 9, and 12 min
- For each subset, data is randomly selected ten times

Training Strategies

From scratch (FS):

- train model from scratch

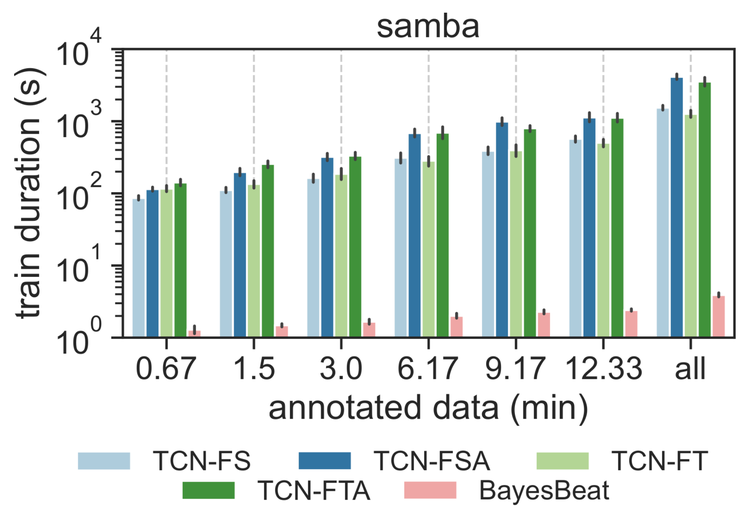
Fine-tuning (FT):

- transfer learning from Western music
- Ballroom, Beatles, GTZAN, and RWC

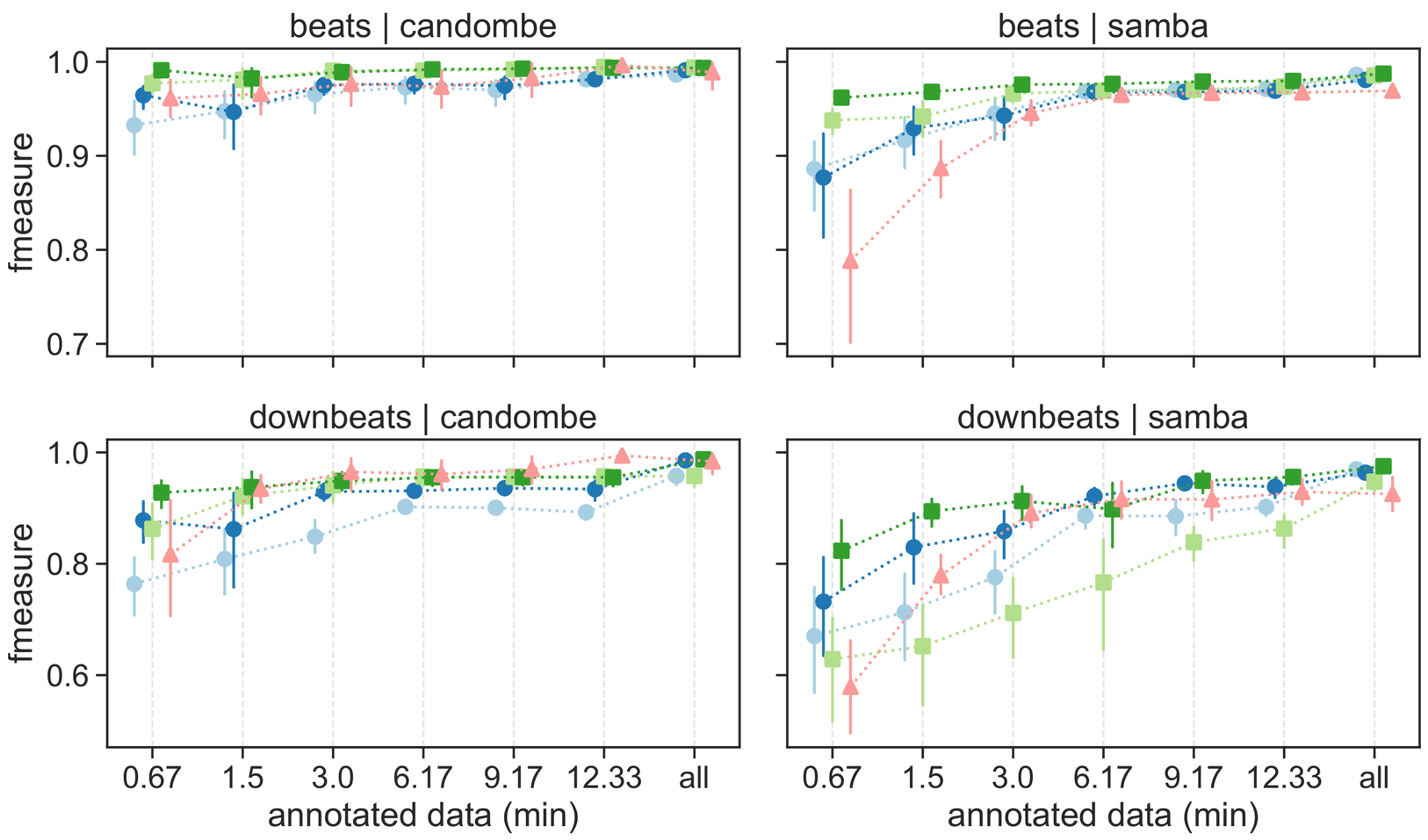
Data Augmentation (FSA/FTA):

- artificially increase training data
- input STFT with varying hop sizes
- repeat strategies FS and FT

Computational Cost



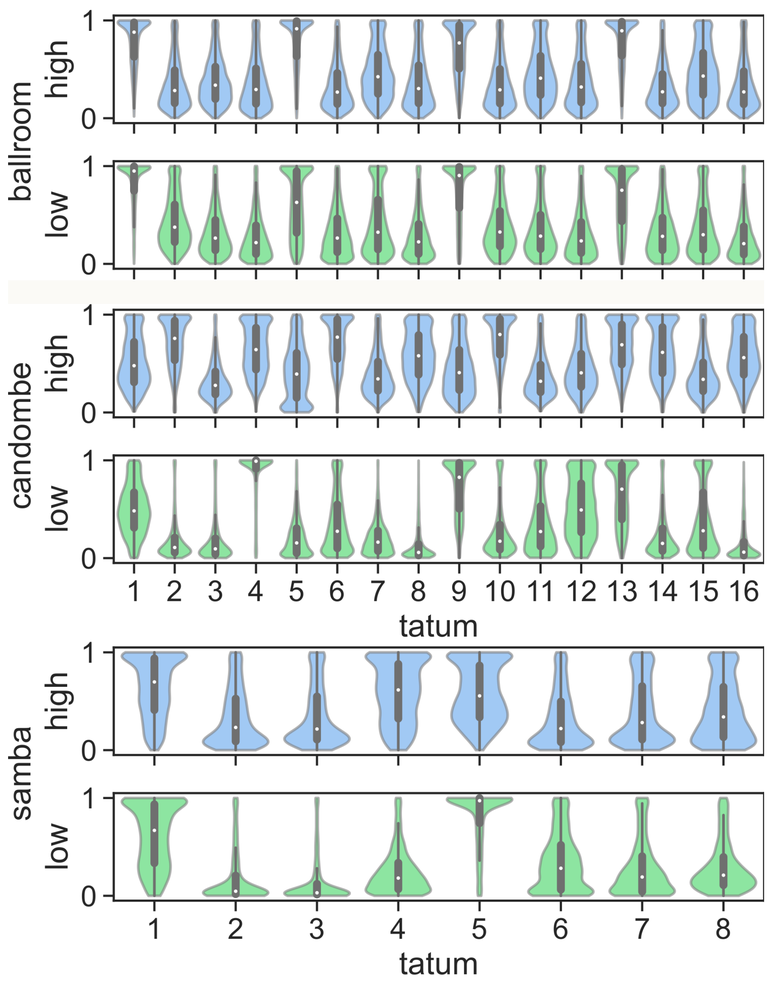
Model Performance



- Good performance with small amount of data
- More data required for *samba* (beat/downbeat)
- Best configuration: FTA; BayesBeat competitive, considering computational cost!

Homogeneity

FS procedure in Ballroom (4/4) yields worse performances in beat (10 to 34%) and downbeat (5 to 20%)



Conclusion

We adapted a meter tracking model using small quantities of data to work in particular Latin American music traditions, *samba* and *candombe*. This seems to be possible only under homogeneity conditions.

Future work: investigate rhythm complexity metrics that could serve to predict the amount of annotated data needed to adapt meter tracking models to particular music genres.

References

[1] S. Böck and M. E. P. Davies. Deconstruct, analyse, reconstruct: How to improve tempo, beat, and downbeat estimation. In *Proc. 21st Int. Soc. Music Inf. Retrieval Conf. (ISMIR)*, pages 574–582, Montreal, Canada, October 2020.

[2] F. Krebs, S. Böck, and G. Widmer. Rhythmic pattern modelling for beat and downbeat tracking from musical audio. In *Proc. 14th Int. Soc. Music Inf. Retrieval Conf. (ISMIR)*, pages 227–232, Curitiba, Brazil, November 2013.

[3] Leonardo Nunes, Martín Rocamora, Luis Jure, and Luiz W. P. Biscainho. Beat and downbeat tracking based on rhythmic patterns applied to the Uruguayan candombe drumming. In *Proc. of the 16th Int. Soc. for Music Information Retrieval Conf. (ISMIR 2015)*, pages 246–270, Málaga, Spain, October 2015.

[4] L. S. Maia, P. D. Tomaz Jr., M. Fuentes, M. Rocamora, L. W. P. Biscainho, M. V. M. Costa, and S. Cohen. A novel dataset of Brazilian rhythmic instruments and some experiments in computational rhythm analysis. In *Proc. 2018 AES Lat. Am. Congr. Audio Eng. (AES LAC)*, pages 53–60, Montevideo, Uruguay, September 2018.