Scale and shift invariant time/frequency representation using auditory statistics: application to rhythm description

Ugo Marchand, Geoffroy Peeters

To cite this version:

Ugo Marchand, Geoffroy Peeters. Scale and shift invariant time/frequency representation using auditory statistics: application to rhythm description. IEEE International Workshop on Machine Learning for Signal Processing, Sep 2016, Vietri Sul Mare, Italy. 2016. <hal-01368888>

HAL Id: hal-01368888
https://hal.archives-ouvertes.fr/hal-01368888

Submitted on 20 Sep 2016
Outline

**Objectives:**
- Creating a representation of the audio signal that differentiates musical rhythms

**Constraints:**
- Creating a representation which is invariant to tempo and to temporal-shifts

**Propositions:**
- Two new 2D (time/frequency) representations of the audio content: 2DMSS and MASSS
- New dataset

**Applications:**
- Use this representation to do auto-tagging, search by similarity

Rhythm description

**Method 2DMSS**

2D Fourier Transform, followed by a 2D Scale Transform known as Fourier-Mellin Transform in Image processing

**Pros/Cons**

Pros: models the relationship between frequency bands and time bins with shift-invariance and scale-invariance
Cons: produces also shift-invariance over frequencies which is an undesired property

**Can differentiate:**

![Diagram showing differentiation between different rhythm types](image)

New dataset: Extended Ballroom

- 698 audio tracks
- 30sec high-quality
- 8 rhythm classes

Extended Ballroom:
- 4,180 audio tracks
- 9+4 rhythm classes
- Similarity annotations

Results

**Classification**

- SVM (MSS, 2DMSS, ccc models)
- Logistic Regression (late-fusion)

**Analysis of results:**

- 2DMSS is not sufficient
- MASSS

-- improves state-of-the-art method by 3% on Ballroom
-- equals state-of-the-art on Cretan dances dataset

**Conclusion:**

- Modeling frequency bands inter-relationship through auditory statistics improves rhythm description

Acknowledgement: This work was funded by the French PIA Bee Music project and the H2020-ICT-2015 ABC DJ project (688122).