Learning Sparse Analytic Filters for Piano Transcription

Frank Cwitkowitz, Mojtaba Heydari, and Zhiyao Duan
This Work

➢ Propose extensions to classic filterbank learning approach

➢ Employ module as frontend to the task of piano transcription
  ○ Replace Mel Spectrogram stage in simple model
Features for MIR

➢ Commonly assume transforms like STFT, CQT, etc., are the best representation of audio for DNNs

○ May not be the case for all MIR tasks
Complex Filterbanks

- These transforms are just complex filterbanks with fixed weights!
- Can represent these with neural networks
  - Fine-tune or learn weights from random initialization

Time Domain Weights
Filterbank Learning (Classic Approach)

- Learn real and imaginary parts independently (or just real)
- Combine into a single magnitude response with $L_2$ pooling
Pitfalls

➢ Learned filters are not analytic

  ○ Not shift-invariant

  ○ Small hop-size required

  ○ Energy at negative frequencies

Frank Cwitkowitz | fcwitkow@ur.rochester.edu
Pitfalls

➢ Learned filters are often noisy
  ○ No localized frequency response
  ○ Very hard to interpret
Proposed Techniques

➢ Only learn the real part of filter and infer imaginary part
  ○ Hilbert transform yields imaginary counterpart to a signal
    ■ Such that the resulting filter is analytic (shift invariant)

➢ Apply variational dropout as regularization to induce sparsity
  ○ Add Gaussian noise with learned variance to response
Experiments

➢ Frontend to Onsets & Frames piano transcription model
➢ Train on MAESTRO and evaluate on MAESTRO/MAPS
➢ Experiment with different variationsinitializations
➢ Conduct an ablation study on proposed techniques
Classic + Random
Hilbert + Random
Hilbert + Random + Variational Dropout
Hilbert + Random + Variational Dropout
Hilbert + Random + Variational Dropout
Hilbert + Random + Variational Dropout
Discussion

➢ Learned filterbanks underperform standard spectral features
➢ Random-initialization on-par with VQT initialization
➢ Lots of interesting observations from filter visualization
Summary

➢ Investigated several variations of a complex filterbank learning module as a frontend for a simple piano transcription model

➢ Techniques to learn analytic filters and to enforce sparsity

All code is available at https://github.com/cwitkowitz/sparse-analytic-filters

Many more filters can be viewed at https://arxiv.org/abs/2108.10382