Comparison of PCA- and Autoencoder-Based Dimensionality Reduction of Feature Sequences for Efficient Music Retrieval

Frank Zalkow, Meinard Müller

International Audio Laboratories Erlangen
Application: Audio Matching for Classical Music

- Application: Identifying a classical piece of music by an audio recording, possibly incomplete
- Given: A database that contains this piece of music, but possibly in a different performance
- Challenges: Differences in tempo, instrumentation, articulation, etc.

Diagram:
- Database
- Query
- Retrieval Procedure
- Ranked List
- Identify Piece of Music
Previous Work

- Comparison of query and database on the basis of chroma features
Previous Work

- Ideal: Subsequence DTW, but prohibitive runtime for large databases
- Previous Work [1]: Shingle approach
- Turned out to be suited: 240 dimensional CENS shingles $\rightarrow$ the basis for our work

### Data Set

<table>
<thead>
<tr>
<th>Composer</th>
<th>Work</th>
<th>Movement</th>
<th>Recording</th>
<th>hh:mm:ss</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beethoven</td>
<td>Op. 67</td>
<td>1</td>
<td>10</td>
<td>01:12:07</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>10</td>
<td>01:44:53</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>10</td>
<td>01:02:53</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>10</td>
<td>01:48:00</td>
</tr>
<tr>
<td>Chopin</td>
<td>Op. 17 No. 4</td>
<td>64</td>
<td></td>
<td>04:36:58</td>
</tr>
<tr>
<td></td>
<td>Op. 24 No. 2</td>
<td>64</td>
<td></td>
<td>02:26:38</td>
</tr>
<tr>
<td></td>
<td>Op. 30 No. 2</td>
<td>34</td>
<td></td>
<td>00:48:11</td>
</tr>
<tr>
<td></td>
<td>Op. 63 No. 3</td>
<td>88</td>
<td></td>
<td>03:09:08</td>
</tr>
<tr>
<td></td>
<td>Op. 68 No. 3</td>
<td>51</td>
<td></td>
<td>01:25:58</td>
</tr>
<tr>
<td>Vivaldi</td>
<td>RV 315</td>
<td>1</td>
<td>7</td>
<td>00:37:40</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>7</td>
<td>00:17:23</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>7</td>
<td>00:20:40</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>362</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>19:30:28</td>
</tr>
</tbody>
</table>
Our Approach

- Our goal: Reducing the dimensionality of the shingles while keeping good matching results
- PCA
- Fully-connected autoencoder
- Fully-convolutional autoencoder
Fully-Connected Autoencoder

![Autoencoder Diagram](image-url)
Fully-Convolutional Autoencoder

32 3x3 Conv
Relu
Maxpool

8 3x3 Conv
Relu
Maxpool

2 3x3 Conv
Relu
Maxpool

1 3x3 Conv
Relu

UpSample
8 3x3 Conv
Relu

UpSample
32 3x3 Conv
Relu

UpSample
1 3x3 Conv
Relu
Experiments

- **PCA**
  - Number of components
  - Preprocessing (standardizing)

- **Deep autoencoder**
  - Size of encoding
  - Preprocessing (standardizing)
  - Fully-connected and fully-convolutional
  - Loss function
  - For convolutional: Ordering of chroma bins in cycle of fifth

- **Evaluation**
  - MAP
  - P@2
  - P@rel
Evaluation

- Comparison of 12-dimensional versions, no standardizing
Future Work

- Different types of regularization
- Variational autoencoder
- Siamese networks
- Using a bigger data set
Evaluation

- No dimensionality reduction

![Graph showing evaluation results with no dimensionality reduction.](image)
Evaluation

- PCA


Evaluation

- Fully-connected network

![Graph showing evaluation results for different network architectures and loss functions.](image-url)
Evaluation

- Fully-convolutional network
Evaluation

- Shallow autoencoder with PCA initialization