A comparison of similarity measures for musical pattern matching

Berit Janssen, Peter van Kranenburg
Thanks to

Tunes & Tales, Meertens Institute

Music Cognition Group, Amsterdam
The study
The study

• Compared six similarity measures
The study

• Compared six similarity measures
• Thresholds
The study

• Compared six similarity measures
• Thresholds
• Pattern length
Results
Results

• String comparison outperforms difference measures
Results

• String comparison outperforms difference measures

• For short-medium patterns, simple patterns perform equally well as more involved measures
Results

• String comparison outperforms difference measures

• For short-medium patterns, simple patterns perform equally well as more involved measures

• Measures which view melodies as curves only become successful for patterns $\geq 6$ notes
Motivation
Studying melodic stability
Studying relationships between variants
Studying relationships between variants

Het was op een Zondagmorgen / Toen kwam hij mij [...]  
Coll Bakker ([1900 ca.]), 057 [nr. 57]

Daar waren twee koningskind'ren / Die hadden [...]  
Coll Bakker ([1900 ca.]), 140 [nr. 140]

Het waren twee koninghs kindren, / sy hadden [...]  
43. Het waren twee koninghs kindren. B.  
Van Duyse (1903-1908), I, p234 [nr. 107]

Het waren twee conincskinderen, / sy hadden [...]  
43. Het waren twee conincskinderen. C.  
Van Duyse (1903-1908), I, p235 [nr. 108]
Use of characteristic licks and patterns
Studying relationships between soloists
Outline

• Similarity measures
• Material
• Music representation
• Pattern matching
• Evaluation method
• Comparison of similarity measures
• Influence of pattern length
Outline

- Similarity measures
- Material
- Music representation
- Pattern matching
- Evaluation method
- Comparison of similarity measures
- Influence of pattern length
Similarity measures

- alignment measures
- simple measures
- curve measures
Similarity measures

• events $i$ in melodies $x$ and $y$
Similarity measures

Levenshtein distance

\[ LD(x_i, y_i) = \min \begin{cases} 
LD(x_{i-1}, y_i) + 1 \\
LD(x_i, y_{i-1}) + 1 \\
LD(x_{i-1}, y_{i-1}) + 1(x_i \neq y_i) 
\end{cases} \]

- ☐ Transposition invariant
- ✔ Time scale invariant
Similarity measures

Substitution distance

\[
SD(x_i, y_i) = \min \begin{cases} 
SD(x_{i-1}, y_i) + 1 \\
SD(x_i, y_{i-1}) + 1 \\
SD(x_{i-1}, y_{i-1}) + |x_i - y_i| 
\end{cases}
\]

- Transposition invariant
- Time scale invariant
Similarity measures

kMismatch

\[ sim_{kmm}(x, y) = \sum_{i=1}^{i=n} 1(x_i \neq y_i) \]

- Transposition invariant
- Time scale invariant
Similitude measures

Difference

\[
sim_{diff}(x, y) = \sum_{i=1}^{i=n} |x_i - y_i|
\]

- Transposition invariant
- Time scale invariant
Similiarity measures

Correlation

\[ sim_{corr}(x, y) = \frac{1}{n} \sum_{i=1}^{i=n} \frac{(x_i - \bar{x})(y_i - \bar{y})}{\sigma_x \sigma_y} \]

- Transposition invariant
- Time scale invariant
4. METHODS

4.1 Matching Process

We extract the pitch values from the original kern files using a sequence of adjacent matches. If the similarity value between the query and a match candidate exceeds a given threshold, the match candidate is retained as a result. For all tested similarity measures, a candidate exceeds a given threshold, the match candidate is registered as a hit when the matched patterns result from a query annotated with those of the matched patterns, and matched with the six similarity measures at various thresholds. We calculated precision and recall scores for the patterns in our comparison which are not transposition invariant. Not all songs of a tune family were notated in the same metric, enabling us to include similarity measures transpose all melodies notated in other tonalities to the G major, which is a common practice in folk song research. We annotated the folk songs as pitch contours, or lists of pitch values, extracted at evenly spaced points in the score. The start of the first melody in Figure 1 is displayed as a dashed, purple line, is compared at each position in the melody using one of the similarity measures. We show the relationship between the derivatives of two B-Spline functions. The time interval for a pattern and the matched pattern is computed by interpolating between data points and is computed by interpolating between data points.

4.2 Implementation of similarity measures

A motif class is annotated. A visualization of the matching in this way to every melody of the tune family in which the annotated motifs and the folk songs are compared is seen in Figure 2. We derive and integrate a sequence of adjacent matches. Moreover, we use a different method for the computation of the area between them up into even shorter segments. We decide to omit this step, as most of our patterns are too short to split shorter segments for subsequent alignment. We use a composite trapezoidal rule, as implemented in scipy.integrate.trapz. Note that our implementation differs from Urbano's method as it does not split the patterns into interpolated patterns. Pitch Derivative is transposition invariant, but in our current implementation, it is sensitive to time scale.

4.3 Music representation

We represent the folk songs as pitch contours, or lists of pitch values, extracted at evenly spaced points in the score. The vast majority of the folk songs in the Dutch Folk Song Database have been notated in the tonality of G major, which is a common practice in folk song research. For each of the duration units, the sounding pitch is represented by MIDI note numbers. For example, if the duration unit of a song is a semiquaver, a crotchet g' for this song would be represented as [67, 67, 67, 67]. Also see Figure 2 for an impression of the resulting pitch contours.

4.4 Evaluation procedure

We calculate precision and recall scores for the patterns matched with the six similarity measures at various thresholds. We show the relationship between the derivatives of two B-Spline functions.

Simularity measures

Pitch derivative

\[ \text{sim}_{pd}(x, y) = \int |x'(t) - y'(t)| dt \]

- Transposition invariant
- Time scale invariant
Outline

• Similarity measures
• Material
• Music representation
• Pattern matching
• Evaluation method
• Comparison of similarity measures
• Influence of pattern length
Material

• Dutch folk song database
• www.liederenbank.nl/mtc
• Annotated corpus: 360 songs
• 1651 annotated motifs, 97 motif classes
Material

NLB070801_01
\[\text{1: gfed} \quad \text{2: aaaa}\]

NLB072154_01
\[\text{1: gfed} \quad \text{2: aaaa}\]
Outline

• Similarity measures
• Material
• **Music representation**
• Pattern matching
• Evaluation method
• Comparison of similarity measures
• Influence of pattern length
Music representation

\[ \text{\textsf{Music engraving by LilyPond 2.14.0—www.lilypond.org}} \]
Outline

• Similarity measures
• Material
• Music representation
• Pattern matching
• Evaluation method
• Comparison of similarity measures
• Influence of pattern length
Pattern matching
Pattern matching

Comparison Step 4

MIDI note number

Duration units
Pattern matching

Comparison Step 8
Outline

• Similarity measures
• Material
• Music representation
• Pattern matching
• Evaluation method
• Comparison of similarity measures
• Influence of pattern length
Evaluation method

Annotated start position

Score

0 0.33 0.66 1

Metrical units from start position

-3 -2 -1 0 1 2 3
Evaluation method

**true positives (tp):**
sum of matching scores

**false positives (fp):**
number of matches, minus tp

**false negatives (fn):**
squared number of matched motifs per motif class, minus tp
Evaluation method

**Precision:**  \[ P = \frac{tp}{tp + fp} \]

**Recall:**  \[ R = \frac{tp}{tp + fn} \]

**F1, F2, F.5**  \[ F1 = 2 \cdot \frac{P \cdot R}{P + R} \]
Outline

• Similarity measures
• Material
• Music representation
• Pattern matching
• Evaluation method
• Comparison of similarity measures
• Influence of pattern length
## Comparison of measures

<table>
<thead>
<tr>
<th>Threshold</th>
<th>Precision</th>
<th>Recall</th>
<th>F.5</th>
<th>F1</th>
<th>F2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.24</td>
<td>0.634</td>
<td>0.527</td>
<td>0.609</td>
<td>0.575</td>
<td></td>
</tr>
<tr>
<td>0.48</td>
<td>0.393</td>
<td>0.687</td>
<td></td>
<td></td>
<td>0.597</td>
</tr>
</tbody>
</table>

(a) Levenshtein Distance

<table>
<thead>
<tr>
<th>Threshold</th>
<th>Precision</th>
<th>Recall</th>
<th>F.5</th>
<th>F1</th>
<th>F2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.24</td>
<td>0.663</td>
<td>0.439</td>
<td>0.602</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.48</td>
<td>0.582</td>
<td>0.557</td>
<td></td>
<td>0.569</td>
<td>0.562</td>
</tr>
</tbody>
</table>

(b) Substitution Distance
### Comparison of measures

<table>
<thead>
<tr>
<th>Threshold</th>
<th>Precision</th>
<th>Recall</th>
<th>F.5</th>
<th>F1</th>
<th>F2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.24</td>
<td>0.633</td>
<td>0.524</td>
<td>0.607</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.42</td>
<td>0.404</td>
<td>0.668</td>
<td>0.504</td>
<td>0.591</td>
<td></td>
</tr>
</tbody>
</table>

#### (c) kMismatch

<table>
<thead>
<tr>
<th>Threshold</th>
<th>Precision</th>
<th>Recall</th>
<th>F.5</th>
<th>F1</th>
<th>F2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>0.656</td>
<td>0.415</td>
<td>0.587</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.46</td>
<td>0.606</td>
<td>0.484</td>
<td></td>
<td>0.538</td>
<td></td>
</tr>
<tr>
<td>0.95</td>
<td>0.419</td>
<td>0.601</td>
<td></td>
<td></td>
<td>0.553</td>
</tr>
</tbody>
</table>

#### (d) Difference
Comparison of measures

<table>
<thead>
<tr>
<th>Threshold</th>
<th>Precision</th>
<th>Recall</th>
<th>F.5</th>
<th>F1</th>
<th>F2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.13</td>
<td>0.327</td>
<td>0.395</td>
<td>0.338</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.24</td>
<td>0.306</td>
<td>0.443</td>
<td>0.362</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.42</td>
<td>0.215</td>
<td>0.541</td>
<td></td>
<td></td>
<td>0.415</td>
</tr>
</tbody>
</table>

(e) Pitch Derivative

<table>
<thead>
<tr>
<th>Threshold</th>
<th>Precision</th>
<th>Recall</th>
<th>F.5</th>
<th>F1</th>
<th>F2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.64</td>
<td>0.140</td>
<td>0.685</td>
<td></td>
<td></td>
<td>0.384</td>
</tr>
<tr>
<td>0.991</td>
<td>0.219</td>
<td>0.393</td>
<td></td>
<td>0.282</td>
<td></td>
</tr>
<tr>
<td>0.999</td>
<td>0.222</td>
<td>0.378</td>
<td>0.280</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Influence of pattern length
Influence of pattern length

• short patterns: \( \leq 2 \) notes
• medium patterns: 3-5 notes
• long patterns: \( \geq 6 \) notes
Influence of pattern length

Levenshtein

Similarity Threshold

Short: 0.534
Medium: 0.646
Long: 0.737
Influence of pattern length

![Graph showing the relationship between similarity threshold and kMismatch for different pattern lengths: Short (0.535), Medium (0.645), and Long (0.724).]
Influence of pattern length

Pitch Derivative

Similarity Threshold

Short 0.279
Medium 0.431
Long 0.682
Influence of pattern length

Correlation

Similarity Threshold

Short

Medium

Long

0.204

0.35

0.58
Results
Results

• String comparison outperforms difference measures
Results

• String comparison outperforms difference measures

• For short-medium patterns, simple patterns perform equally well as more involved measures
Results

• String comparison outperforms difference measures

• For short-medium patterns, simple patterns perform equally well as more involved measures

• Measures which view melodies as curves only become successful for patterns $\geq 6$ notes
Perspectives
Perspectives

• Influence of music representation
Perspectives

• Influence of music representation
• Different pattern annotations
Perspectives

• Influence of music representation
• Different pattern annotations
• Different measures
Perspectives

• Influence of music representation
• Different pattern annotations
• Different measures
• Efficient implementations
Thank you!