Search-Effectiveness Measures for Symbolic Music Queries in Very Large Databases

Craig Stuart Sapp
craig@ccrma.stanford.edu

Yi-Wen Liu
jacobliu@stanford.edu

Eleanor Selfridge-Field
esfield@stanford.edu

ISMIR 2004
Universitat Pompeu Fabra
Barcelona, Spain
12 October 2004
Introduction

Match-Count Profiles
Musical Features

• We examined search characteristics of 14 musical features:

7 Pitch features
(examples below) + 7 Rhythm features: (3 duration & 4 metric)

1. duration (37-74)
2. duration gross contour (3)
3. duration refined contour (5)
4. beat level (2)
5. metric level (10-14)
6. metric gross contour (3)
7. metric refined contour (5)

• How do all these different features affect searching in a database?
Anchored vs. Unanchored Searches

**Search Pattern:** F A C

Two types of search methods, Examples:

**Anchored Matches**

search only from the start of a database entries

**Unanchored Matches**

search starting at any position in database entries
## Example Feature Searches

![Musical Example](image)

<table>
<thead>
<tr>
<th>Feature</th>
<th>Query (in Themefinder)</th>
<th>Anchored Matches</th>
<th>Unanchored Matches</th>
</tr>
</thead>
<tbody>
<tr>
<td>pitch name</td>
<td>pch</td>
<td>464</td>
<td>1,710</td>
</tr>
<tr>
<td>12-tone pitch</td>
<td>12p</td>
<td>464</td>
<td>1,710</td>
</tr>
<tr>
<td>musical interval</td>
<td>mi</td>
<td>+M3 +m3</td>
<td>1,924</td>
</tr>
<tr>
<td>12-tone interval</td>
<td>12i</td>
<td>+4 +3</td>
<td>1,925</td>
</tr>
<tr>
<td>scale degree</td>
<td>sd</td>
<td>1 3 5</td>
<td>2,009</td>
</tr>
<tr>
<td>pitch refined contour</td>
<td>prc</td>
<td>U U</td>
<td>4,677</td>
</tr>
<tr>
<td>pitch gross contour</td>
<td>pgc</td>
<td>U U</td>
<td>19,787</td>
</tr>
</tbody>
</table>

Searching a database of 100,000 melodic incipits/themes
Raw Data Extraction

- Now plot measurements as a “match-count profile”
  
  - **x-axis:** query length
  - **y-axis:** match count (log scale)
Individual Match-Count Profile

12-tone interval features:

- Anchored and Unanchored searches merge at length = 8
- Unique match found at length = 10
Interesting Query Lengths

Target incipit:

TTU = length of query yielding unique match
TTS = length giving matches under limit size

How long query length must be to generate a sufficiently small set of matches
e.g., first search-length which gives fewer than 10 matches
Average Match-Count Profiles

- Average all target profiles over entire database:
Average Match-Count Profiles

• Average all target profiles over entire database:
Anchored/Unanchored Profile Slopes

- Anchored Searching: $O(\log N)$
- Unanchored Searching: $O(N^2)$

- Anchored/Unanchored slopes not much different.
- Anchored searching is much faster.
Match-Count Profiles for Pitch Features

- Steeper initial slope = more descriptive feature

- Twelve-tone pitch and full pitch spelling features are very identical (orange curve)
- Absolute twelve tone pitch and relative twelve-tone interval are close.
- 7-symbol scale degree features close to 5-symbol refined pitch contour.
- 3-symbol pitch gross contour more descriptive than 3-symbol duration gross contour.
- TTS for rhythm twice as long than pitch TTS.
- TTS for gross metric descriptions 5 times as long as pitch TTS values.
- Rhythm feature curves more crooked.

Phrase/meter effects?
Four Applications of Profiles:

- Entropy & Entropy Rate
- Joint Feature Analysis
- Match Count Predictions
- Synthetic Database Analysis
Entropy

- Entropy measures basic information content of a musical feature also called “Shannon Entropy” or “First-order Entropy”

\[ H(X) \triangleq - \sum_i P_i(X) \log_2 P_i(X) \]

Entropy (bits/symbol) Normalized probability distribution

Example calculation:

\[ H(X) = - \sum_i P_i(X) \log_2 P_i(X) \]

\[ H(12i) = 3.41165 \text{ bits/pitch} \]

\[ \rightarrow 3.4 \text{ bits/note is the minimum symbol storage size needed to store sequences of 12-tone intervals (Folksong data set).} \]
Entropy Rate

- Entropy is a contextless (memoryless) measure.
- *Real music features are related to surrounding musical context.*
- Average entropy (entropy-rate) is more informative:

\[
G(N) \triangleq \frac{H(X^N)}{N}
\]

“Nth-order” entropy

also called

“Average Entropy”

Entropy rate (bits/symbol)

N=Sequence length

Entropy & entropy rate for various repertories:

Note:

\[
G(N) \leq H(X)
\]
Entropy-Rate Estimation from TTS

- *Entropy* characterizes the minimum possible average TTS.
- *Entropy-rate* characterizes the actual average TTS.

\[
G = \frac{\log_2 \left( \frac{M}{k} \right)}{\text{TTS}_k}
\]
Applications of Profiles

- Entropy & Entropy Rate
- Joint Feature Analysis
- Match Count Predictions
- Synthetic Database Analysis
Joint Feature Analysis

• How independent/dependent are pitch and rhythm features?

• What is the effect of searching pitch and rhythm features in parallel?
Mutual Information

- Measurement of the correlation of two types of features

$$H(a)$$

- e.g., pitch

$$H(b)$$

- e.g., rhythm

Joint entropy

$$H(a, b)$$

$$I(a; b) = H(a) + H(b) - H(a, b)$$

Conditional entropy

$$H(a|b) = H(a, b) - H(b)$$

$$H(b|a) = H(a, b) - H(a)$$

Mutual information
Combining Pitch and Rhythm Searches

\[
\begin{align*}
H(pgc) &= 1.5325 \\
H(rgc) &= 1.4643 \\
H(pgc, rgc) &= 2.9900 \\
I(pgc; rgc) &= H(pgc) + H(rgc) - H(pgc, rgc) = 0.0068
\end{align*}
\]

- Pitch and Rhythm are very independent features.
  (at least for \textit{pgc+rgc} averaged over entire database)
- Therefore, combining independent search features should be effective.
Joint Feature Profiles
for pgc/rgc vs. twelve-tone interval searching

- 3*3 states work as well as 88 twelve-tone interval states.
- \textit{pgc} and \textit{rgc} are generic features less prone to query errors.
Joint Feature Search Effectiveness

Single Feature Match Count Profiles

Joint Feature Match Count Profiles
Joint Feature Search Effectiveness

Single Feature Match Count Profiles

Log$_2$ matches

Joint Feature Match Count Profiles

Log$_2$ matches

TTS: 6 - 11

(All dataset)
Applications of Profiles

• Entropy & Entropy Rate
• Joint Feature Analysis
• Match Count Predictions
• Synthetic Database Analysis
Expectation Function

• Entropy Rate can be used to predict the number of matches:

\[ E(n) = \frac{M}{R^n} \]

\[ R = 2^H \]

(H = measured entropy rate)

Expected match counts for an \( n \)-length query

• Example:

• Consider a database of “best 3 out of 5” Heads/Tails coin flips:

| H H T H T | T H T T H | H T T H H | T T T H H | H H H H H |

Entropy Rate = Entropy = \( \log_2 2 = 1 \) bit/symbol

Therefore \( R = 2^{\log_2 2} = 2^1 = 2 \)

• Likelyhood starting sequence is “H”: 50% \( \Rightarrow E(1) = \frac{M}{2^1} = \frac{M}{2} \)

• Likelyhood starting sequence is “H T”: 25% \( \Rightarrow E(2) = \frac{M}{2^2} = \frac{M}{4} \)

• Likelyhood starting sequence is “H H ”: 25% \( \Rightarrow E(2) = \frac{M}{2^2} = \frac{M}{4} \)
Match-Count Profile Constraint

- The match-count profile queries are constructed from database entries.
- Therefore at least one match is always expected.
- Steal this guaranteed match from M, and add as a constant to the expectation function:

\[ E(n) = \frac{M - 1}{R^n} + 1 \]

- How to get rid of curvature caused by constant +1 term?
Match-Count and Derivative Profile Comparison

Match-Count Profile expectation function:

\[ E(n) = \frac{M - 1}{R^n} + 1 \]

To measure the entropy rate of small databases, you would need to use the derivative plot since the +1 term would be too powerful.

Initial slope of both profiles is the entropy rate.

Derivative profile maintains entropy-rate slope for larger query lengths.

What about \( E(n) - 1 \)?
Expectation Plot Functions

“Match-Count Profile”

“Derivative Profile”

“Target-Exclusion Profile”

- Removes +1 curvature and *not* sensitive to duplicate entries in the database.
- Best method for measuring entropy-rate

- Removes +1 curvature, but sensitive to duplicate entries in the database

Using unanchored rhythmic features for Essen-full
Applications of Profiles

- Entropy & Entropy Rate
- Joint Feature Analysis
- Match Count Predictions
- Synthetic Database Analysis
Synthetic vs. Real Database Profiles

Legend:

- **Uniform random data**
- **Weighted Random** Based on real data probability distribution.
- **Markov process generated data**
- **Real data**

Synthetic databases will not curve.
Effects of Duplicate Entries on Profiles

Duplicate entries in the database do not have a significant effect on entropy-rate measurements:

- $E(n)$ and $E(n)-1$ profiles can be used to measure amount of duplication in database.

- $E(n) - E(n+1)$ removes effect of duplicate entries entirely.

E(n) - E(n+1) is very sensitive to detecting duplication.
Effect of Incipit Length on Profiles

- An incipit a short initial excerpt from a full composition
- How short is is too short for a musical incipit?

Derivative Profile

slope at long query lengths is artificially increased when incipits are too short.

shorter incipits cause quantization noise in low match-count region
Search-Effectiveness Measures for Symbolic Music Queries in Very Large Databases

Craig Stuart Sapp  
craig@ccrma.stanford.edu

Yi-Wen Liu  
jacobliu@stanford.edu

Eleanor Selfridge-Field  
esfield@stanford.edu

ISMIR 2004  
Universitat Pompeu Fabra  
Barcelona, Spain  
12 October 2004
Interesting metrics for analyzing the effectiveness of search features:

• **Match-Count Profiles:** Examines match characteristics of a musical feature for longer and longer queries.

• **Entropy Rate:** Characterizes match count profiles well with a single number. Useful for predicting the expected average number of matches for a given length query.

• **TTS:** The number of symbols in query necessary to generate a sufficiently small number of matches (average). TTU not as useful due to noise.
Proof for Derivative Plots

\[ E(n) = \frac{M - 1}{R^n} + 1 \]  
(expectation function for Match-Count Profiles)

\[ E(n) - E(n + 1) = \frac{M - 1}{R^n} - \frac{M - 1}{R^{n+1}} \]  
(subtract \( n \) and \( n+1 \) values of \( E(\ ) \) to cancel +1 term)

\[ E(n) - E(n + 1) = \frac{(R - 1)(M - 1)}{R R^n} \]  
(algebra manipulation)

plotting on a log scale, so take the log of both sides:

\[ \log_2[E(n) - E(n + 1)] = \log_2 \left[ \frac{(R - 1)(M - 1)}{R} \right] - \log_2 R^n \]

Let: \( y = \log_2[E(n) - E(n + 1)] \) and \( b = \log_2 \left[ \frac{(R - 1)(M - 1)}{R} \right] \)

so the equation becomes:

\[ y = b - \log_2 R^n \]

since \( R = 2^H \)

Let: \( x = n \)

\[ y = -Hx + b \]

which is a line with a slope proportional to the entropy (rate)
Derivative Plots for 12i features

- Vocal music tending to lower entropy rates
- Luxembourg set has most predictable interval sequences.
- Latin Motets (vocal) have highest entropy-rate for twelve-tone intervals.
Themefinder Website

http://www.themefinder.org
# Themefinder Collections

<table>
<thead>
<tr>
<th>Data set</th>
<th>Count</th>
<th>Web Interface</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classical</td>
<td>10,718</td>
<td>themefinder.org</td>
</tr>
<tr>
<td>Folksong</td>
<td>8,473</td>
<td>themefinder.org</td>
</tr>
<tr>
<td>Renaissance</td>
<td>18,946</td>
<td>latinmotet.themefinder.org</td>
</tr>
<tr>
<td>US RISM A/II</td>
<td>55,490</td>
<td></td>
</tr>
<tr>
<td>Polish</td>
<td>6,060</td>
<td></td>
</tr>
<tr>
<td>Luxembourg</td>
<td>612</td>
<td>lux.themefinder.org</td>
</tr>
</tbody>
</table>

*total: 100,299*
Matches on First Seven Notes

A.

B.

C.

D.

E.

F. x2

G.

H. x4
Entropy and Entropy Rate
for various repertories in the Themefinder database

\[ G(N) \leq H(X) \]
Entropy rate less than or equal to the Entropy
Search Failure Rates

Database size: 100,299
Average note count/incipit: 16

- Plot measures how often a search produces too many matches for query sequences as long as the database entry.
**Time To Uniqueness**

TTU = the number of query symbols needed to find the exact match in the database. Turns out to not be very useful since it is more susceptible to noise in the data.
Effect of Incipit Length on Profiles

Derivative Curve

- Shorter incipits cause quantization noise in low match-count region.
- Slope at long query lengths is artificially increased when incipits are too short.

Effect of Incipit Length on Anchored Match Profile

- Shorter incipits cause quantization noise in low match-count region.
- Red line = full song (51 notes average)
- Black line = incipit (17 notes average)
- Greying lines each remove one note from incipit.
3.4 bits/note is the lower symbol storage size limit needed to store sequences of 12-tone intervals (Folksong data set).

- Entropy can be used as a basic estimate for how many notes are necessary to find a unique/sufficient match in the database, but ...
Expectation Function

\[ M \quad = \quad \text{database size} \]
\[ E(n) \quad = \quad \text{average expected match counts for an } n\text{-length query} \]
\[ R = 2^H \quad \text{where } H \text{ is the entropy rate of the feature being searched for} \]
\[ \quad \text{(Entropy rate is assumed to be constant)} \]

In general: \[ E(n) = \frac{M}{R^n} \]

For example, consider sequences created with a uniform random distribution of three states (the next symbol in the sequence is equally likely to be any of the three states). Then, the entropy of the sequence is: \[ H = \log_2 3 \] which makes \[ R = 2^{\log_2 3} = 3 \]

and the formula for the expected match counts becomes: \[ E(n) = \frac{M}{3^n} \]

then 1/3 of the database entries should be matched with a one-length query on the average:
\[ E(1) = \frac{M}{3^1} = \frac{M}{3} \]

and a length-two query should return 1/9 of the database on the average:
\[ E(2) = \frac{M}{3^2} = \frac{M}{9} \]
Joint Pitch/Rhythm Effects on TTS

- Adding $rgc$ to pitch features usually reduces the search length by 2 notes.
- Combining $rgc$ and $pgc$ reduces search length by 4 notes.

Chinese Folksongs dataset

Classical dataset