

Fingerprint 1:

	1	2	3	4	5	6	7	8	9
w	[-2.5]	[1.0, 1.5]	[2.5]	[5.0]	[-2.5]	[4.5]	[4.0]	[4.5]	[-2.5]
0.0	0	0	0	0	0	0	0	0	0
1.0	-1.5	-1	-1.5	4	-3.5	-0.5	-1	5	NaN
2.0	-2.5	-2.5	2.5	0.5	-4	-1.5	4	NaN	NaN
3.0	-4	1.5	-1	0	-5	3.5	NaN	NaN	NaN
4.0	0	1.5	5	0	NaN	NaN	NaN	NaN	NaN
5.0	-3.5	-2.5	-2.5	4	NaN	NaN	NaN	NaN	NaN
6.0	5	5	0	0	NaN	NaN	NaN	NaN	NaN
7.0	-5	1.5	Nan	Nan	Nan	Nan	Nan	Nan	Nan
8.0	0	Nan	Nan	Nan	Nan	Nan	Nan	Nan	Nan

# The Fingerprint Algorithm: Detecting and Quantifying Similarity in Fiddle Tunes

Fingerprint 2:

	1	2	3	4	5	6	7	8	9
w	[-2.5]	[1.5]	[0.0]	[1.5]	[-2.5]	[1.5]	[-0.5]	[-1.0]	[-2.5]
0.0	0	0	0	0	0	0	0	0	0
1.0	-1.5	-1	-1.5	4	-4	0.5	-1.5	5	NaN
2.0	-2.5	-2.5	2.5	0	-3.5	-1	3.5	NaN	NaN
3.0	-4	1.5	-1.5	0.5	NaN	NaN	NaN	NaN	NaN
4.0	5	5	0	0	NaN	NaN	NaN	NaN	NaN
5.0	-3.5	-3.5	2.5	Nan	Nan	Nan	Nan	Nan	Nan
6.0	5	5	0	0	NaN	NaN	NaN	NaN	NaN
7.0	-5	1.5	Nan	Nan	Nan	Nan	Nan	Nan	Nan
8.0	0	Nan	Nan	Nan	Nan	Nan	Nan	Nan	Nan



**McGill**



**Schulich School of Music**  
**École de musique Schulich**



BIBLIOTHÈQUE  
ET ARCHIVES  
NATIONALES  
DU QUÉBEC

# Overview

- Background
- Why an algorithm for similarity?
- Why fiddle tunes?
- Existing initiatives
- The fingerprint algorithm
  - What does it do?
  - How does it work?
- Results
  - Comparison of two fingerprints
  - Rankings
- Going forward
  - Creating a visual representation
  - Challenges
  - Next steps

# Background



- In Montreal in the 1920s and 1930s, fiddlers, accordionists and harmonica players recorded over 1000 sides of instrumental dance tunes for commercial labels.
  - These 78s offer the first large-scale musical documentation of a working-class repertoire then popular in the home, on stage and on radio.

Images source: The Virtual Gramophone,  
<http://www.collectionscanada.gc.ca/gramophone/index-e.html>

# Background



- The musical contents of these recordings remain largely unexamined. The only detailed musicological study of this repertoire (Duval 2012) is limited to a subset of metrically irregular tunes.
- This project proposes a computer-aided approach to parsing this repertoire. Specifically, it proposes an algorithm that searches not only for identical but also for similar melodic phrases, and quantifies that similarity.

# Why is a similarity algorithm necessary for this repertoire?

- Frequent melodic and rhythmic variation: certain tunes were recorded multiple times (usually by different artists) and these settings are often quite varied
  - Occasionally we can identify these as variants because they share a common title, as for the tune “Money Musk”
  - However, this is not the case for most variants



# Why is a similarity algorithm necessary for this repertoire?

- Non-standardized titles
  - Many tunes are variants of common Scottish, Irish and American tunes but were not recorded under the usual titles
  - Ex.: three variants of the Scottish song “My Love She’s But A Lassie Yet”
- The only way to identify these as variants is to compare the melodies.



1928



1931



1937

# Why is a similarity algorithm necessary for this repertoire?

- This repertoire is too large for one person to hold all the melodies in his/her head at once. As a result, identifying variants is something like searching for a needle in a haystack.
  - Given a tune, this algorithm will search for tunes that are similar and quantify that similarity.



# Why use a database of fiddle tunes from Quebec?

- Many examples of melodic and rhythmic variation => a diverse set of test cases
  - The many recordings of “Money Musk” (from 1920s to the present) serve as an encyclopedia of variation techniques and may be used to generate comparison and analysis functions.
- Ease of use: the tunes are monophonic, relatively short and often follow standard forms (AB, AABB, ABC, ABCD, etc.)
- Unique incipits: the first two bars of a tune are nearly always enough to identify the tune
- Availability: many of these tunes are accessible via the Virtual Gramophone (Library and Archives Canada). Nearly all have been catalogued by Labbé (1995).



## Existing initiatives

Global Chant Database (<http://globalchant.org>): searches for a given string of notes in a database of over 24,000 chant incipits. Returns exact matches and transpositions occurring at the beginning of a chant, and instances of the interval pattern within melodies.

Early American Secular Music Index (<http://www.colonialdancing.org/Easmes/Index.htm>): Searches for a given string of notes in a database of over 75,000 items. However, there is no actual search engine (to my knowledge). Rather, the user selects an incipit “of interest” or a series of “stress notes” (the first notes of consecutive bars) from a web page listing all of the options.

Scottish Music Index (<http://www.scottishmusicindex.org>): Searches for a numerical theme code in a database of over 12,000 such codes. Each code is a numerical representation (by scale degree) of the strong beats of the first two bars of a tune. The database contains theme codes for most printed Scottish tunes from the eighteenth century to the early twentieth century.

Traditional Tune Archive (<http://tunearch.org/wiki/TTA>): This wiki is currently under construction. It uses the same theme code system as the Scottish Music Index. The search function will not find partial strings (“11H6” will only return tunes in 3/4, not tunes in 4/4 that begin with this string).

## Existing initiatives

TunePal (<http://tunepal.org/tunepal/index.php>) : This is an excellent tool for identifying Irish tunes from audio sources. The program uses the MATT2 tools developed by Bryan Duggan (PhD thesis, 2009). TunePal translates audio into symbolic notation (a simplified setting in ABC) and compares that to a database of over 24,000 Irish tunes, also in simplified ABC notation. The database is “crowd-sourced” from several dozen online sites that contain ABC transcriptions of tunes (mostly Irish tunes).

TunePal uses an edit distance algorithm to compare tunes. This is very effective if the unknown tune is an exact or near-exact match to a tune in the database. However, a variant with significant melodic alterations may not register. In addition, TunePal does not identify transpositions. Note that TunePal looks for exact or similar strings of ABC anywhere in the tune, regardless of metrical placement.

- I have used TunePal to identify several common Irish and American tunes that were recorded in Montreal under new names.

*Note that Duggan (2009:65-78) includes a thorough review of existing (in 2009) methods for detecting melodic similarity.*

# The fingerprint algorithm

# The fingerprint algorithm

Goals:

- Search for likely variants of a given tune while allowing for significant melodic and rhythmic variation
- Describe and quantify the similarities and differences of multiple variants of a given tune

Process:

- Replicate the typical thought processes of musicians attempting to identify variants

# The fingerprint algorithm

Considerations:

- Each tune has two or more strains (A, B, etc., commonly called “parts”)
- The two strains must be treated separately because musicians sometimes borrowed A and B strains from two different tunes in order to create a new tune.
- Metrical placement matters: the notes that fall on the strong beats are more essential to the identity of the strain than those on the weak beats.
- Transposition is common and tonal centers are sometimes ambiguous.
- For nearly all strains, the first two bars (plus one beat) can serve as a unique identifier: a “fingerprint.”

# Eight fingerprints for the first strain of “Money Musk”

Musical notation for the A strain of "Money Musk" from Allard 1928. The music is in G major (one sharp) and common time. It consists of two measures of eighth-note patterns.

Allard 1928, A strain

Musical notation for the A strain of "Money Musk" from Lajoie 1951. The music is in G major (one sharp) and common time. It consists of four measures of eighth-note patterns.

Lajoie 1951, A strain

Musical notation for the A strain of "Money Musk" from Boivin 1974. The music is in G major (one sharp) and common time. It consists of four measures of eighth-note patterns.

Boivin 1974, A strain

Musical notation for the C strain of "Money Musk" from Lajoie 1951. The music is in G major (one sharp) and common time. It consists of four measures of eighth-note patterns.

Lajoie 1951, C strain

Musical notation for the C strain of "Money Musk" from Boivin 1974. The music is in G major (one sharp) and common time. It consists of four measures of eighth-note patterns.

Boivin 1974, C strain

Musical notation for the E strain of "Money Musk" from Lajoie 1951. The music is in G major (one sharp) and common time. It consists of four measures of eighth-note patterns.

Lajoie 1951, E strain

Musical notation for the D strain of "Money Musk" from Duchesne 1934. The music is in G major (one sharp) and common time. It consists of eight measures of eighth-note patterns.

Duchesne 1934, D strain

Musical notation for the G strain of "Money Musk" from Soucy 1927. The music is in G major (one sharp) and common time. It consists of eight measures of eighth-note patterns.

Soucy 1927, G strain

# Eight fingerprints for the first strain of “Money Musk”

*All starting on G for easier comparison*



Musical notation for the A strain of "Money Musk" from Allard 1928. The key signature is one sharp (F#). The melody consists of eighth and sixteenth notes, starting on G.

Allard 1928, A strain



Musical notation for the A strain of "Money Musk" from Lajoie 1951, transposed from C major. The key signature is one sharp (F#). The melody consists of eighth and sixteenth notes, starting on G.

Lajoie 1951, A strain (transposed from C major)



Musical notation for the A strain of "Money Musk" from Boivin 1974. The key signature is one sharp (F#). The melody consists of eighth and sixteenth notes, starting on G.

Boivin 1974, A strain



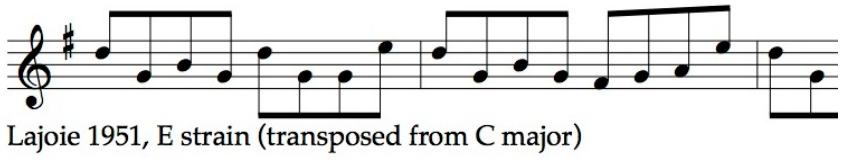
Musical notation for the C strain of "Money Musk" from Lajoie 1951. The key signature is one sharp (F#). The melody consists of eighth and sixteenth notes, starting on G.

Lajoie 1951, C strain



Musical notation for the C strain of "Money Musk" from Boivin 1974. The key signature is one sharp (F#). The melody consists of eighth and sixteenth notes, starting on G.

Boivin 1974, C strain



Musical notation for the E strain of "Money Musk" from Lajoie 1951, transposed from C major. The key signature is one sharp (F#). The melody consists of eighth and sixteenth notes, starting on G.

Lajoie 1951, E strain (transposed from C major)



Musical notation for the D strain of "Money Musk" from Duchesne 1934. The key signature is one sharp (F#). The melody consists of eighth and sixteenth notes, starting on G.

Duchesne 1934, D strain



Musical notation for the G strain of "Money Musk" from Soucy 1927, transposed from A major. The key signature is one sharp (F#). The melody consists of eighth and sixteenth notes, starting on G.

Soucy 1927, G strain (transposed from A major)

## What does the algorithm do?

The algorithm compares the first two bars (the “fingerprints”) of two given strains. This comparison is effected through a series of functions that:

- compare strong beats to strong beats
- establish an optimal correspondence between strong beats
- search for missing strong beats on corresponding weak beats
- compare weak beats to weak beats
- compare contour and check for strong-beat reversal on consecutive non-matching strong beats, and then check for alignment between corresponding weak beats

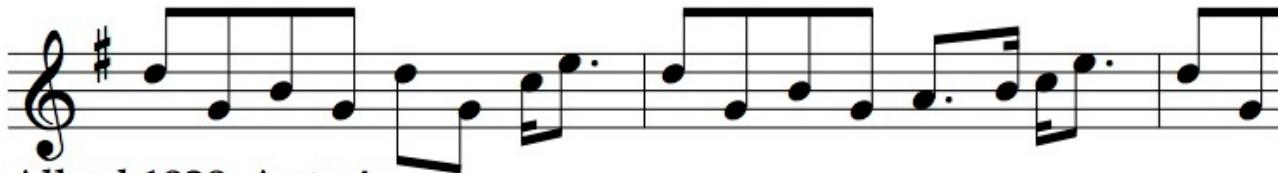
Note that this comparison is interval-based and is therefore independent of tonal center.

# What does the algorithm do?

The algorithm compares the first two bars (the “fingerprints”) of two given strains. This comparison is effected through a series of functions that:

- compare strong beats to strong beats
- establish an optimal correspondence between strong beats
- search for missing strong beats on corresponding weak beats
- compare weak beats to weak beats
- compare contour and check for strong-beat reversal on consecutive non-matching strong beats, and then check for alignment between corresponding weak beats

Ex. 1



Allard 1928, A strain

Musical notation for 'Allard 1928, A strain' in G major (one sharp) and common time. The notation consists of two measures. Measure 1 starts with a quarter note, followed by an eighth note, another eighth note, and a dotted eighth note. Measure 2 starts with a quarter note, followed by an eighth note, a sixteenth note, and a dotted eighth note. The notes are connected by vertical stems.



Boivin 1974, A strain

Musical notation for 'Boivin 1974, A strain' in G major (one sharp) and common time. The notation consists of two measures. Measure 1 starts with a quarter note, followed by an eighth note, another eighth note, and a dotted eighth note. Measure 2 starts with a quarter note, followed by an eighth note, a sixteenth note, and a dotted eighth note. The notes are connected by vertical stems.

# What does the algorithm do?

The algorithm compares the first two bars (the “fingerprints”) of two given strains. This comparison is effected through a series of functions that:

- compare strong beats to strong beats and identify **matched SB** (i.e., same pitch)
- establish an optimal correspondence between strong beats (=> **non-matched SB**)
- search for missing strong beats on corresponding weak beats
- compare weak beats to weak beats
- compare contour and check for strong-beat reversal on consecutive non-matching strong beats, and then check for alignment between corresponding weak beats

Ex. 1

Allard 1928, A strain

Boivin 1974, A strain

# What does the algorithm do?

The algorithm compares the first two bars (the “fingerprints”) of two given strains. This comparison is effected through a series of functions that:

- compare strong beats to strong beats and identify **matched SB** (i.e., same pitch)
- establish an optimal correspondence between strong beats (=> **non-matched SB**)
- search for **missing strong beats on corresponding weak beats**
- compare weak beats to weak beats
- compare contour and check for strong-beat reversal on consecutive non-matching strong beats, and then check for alignment between corresponding weak beats

Ex. 1

Allard 1928, A strain

Boivin 1974, A strain

# What does the algorithm do?

The algorithm compares the first two bars (the “fingerprints”) of two given strains. This comparison is effected through a series of functions that:

- compare strong beats to strong beats and identify **matched SB** (i.e., same pitch)
- establish an optimal correspondence between strong beats (=> **non-matched SB**)
- search for **missing strong beats on corresponding weak beats**
- compare weak beats to weak beats and identify **matched WB on matched SB** and **matched WB on non-matched SB**
- compare contour and check for strong-beat reversal on consecutive non-matching strong beats, and then check for alignment between corresponding weak beats

Ex. 1

The image shows two staves of musical notation. The top staff is labeled "Allard 1928, A strain" and the bottom staff is labeled "Boivin 1974, A strain". Both staves have a treble clef and a key signature of one sharp. The notation consists of eighth and sixteenth notes. Colored arrows indicate correspondences between notes in the two strains. Red double-headed arrows connect notes in the first measure of both strains. In the second measure, there are green double-headed arrows between the first and third notes, and orange double-headed arrows between the second and fourth notes. A purple arrow points from the second note of the Allard staff to the second note of the Boivin staff. The third measure shows red double-headed arrows between the first and third notes, and green double-headed arrows between the second and fourth notes. The fourth measure shows red double-headed arrows between the first and third notes, and orange double-headed arrows between the second and fourth notes.

# What does the algorithm do?

The algorithm compares the first two bars (the “fingerprints”) of two given strains. This comparison is effected through a series of functions that:

- compare strong beats to strong beats and identify matched SB
- establish an optimal correspondence between strong beats (=> non-matched SB)
- search for missing strong beats on corresponding weak beats
- compare weak beats to weak beats and identify matched WB on matched SB and matched WB on non-matched SB
- compare contour and check for strong-beat reversal on consecutive non-matching strong beats, and then check for alignment between corresponding weak beats

Ex. 2



Allard 1928, A strain

Musical notation for 'Allard 1928, A strain' in G major (one sharp). The notation consists of two measures. Measure 1 starts with a quarter note followed by eighth notes. Measure 2 starts with a half note, followed by eighth notes. The notes are connected by vertical stems.



Lajoie 1951, E strain (transposed from C major)

Musical notation for 'Lajoie 1951, E strain (transposed from C major)' in E major (no sharps or flats). The notation consists of two measures. Measure 1 starts with a quarter note followed by eighth notes. Measure 2 starts with a half note, followed by eighth notes. The notes are connected by vertical stems.

# What does the algorithm do?

The algorithm compares the first two bars (the “fingerprints”) of two given strains. This comparison is effected through a series of functions that:

- compare strong beats to strong beats and identify **matched SB**
- establish an optimal correspondence between strong beats ( $\Rightarrow$  **non-matched SB**)
- search for missing strong beats on corresponding weak beats
- compare weak beats to weak beats and identify matched WB on matched SB and matched WB on non-matched SB
- compare contour and check for strong-beat reversal on consecutive non-matching strong beats, and then check for alignment between corresponding weak beats

Ex. 2

The image shows two musical staves. The top staff is labeled "Allard 1928, A strain" and the bottom staff is labeled "Lajoie 1951, E strain (transposed from C major)". Both staves are in treble clef and have a key signature of one sharp. The top staff has a tempo marking of 120 BPM. Red double-headed arrows indicate strong beat correspondences between the two strains. Orange double-headed arrows indicate non-corresponding strong beats where a weak beat in one strain aligns with a strong beat in the other. The music consists of eighth and sixteenth note patterns.

# What does the algorithm do?

The algorithm compares the first two bars (the “fingerprints”) of two given strains. This comparison is effected through a series of functions that:

- compare strong beats to strong beats and identify **matched SB**
- establish an optimal correspondence between strong beats (=> **non-matched SB**)
- search for **missing strong beats on corresponding weak beats**
- compare weak beats to weak beats and identify matched WB on matched SB and matched WB on non-matched SB
- compare contour and check for strong-beat reversal on consecutive non-matching strong beats, and then check for alignment between corresponding weak beats

Ex. 2

The image shows two musical staves, each consisting of five horizontal lines. The top staff is labeled "Allard 1928, A strain" and the bottom staff is labeled "Lajoie 1951, E strain (transposed from C major)". Both staves begin with a treble clef and a key signature of one sharp (F#). The music consists of two measures. In the first measure, there are four eighth notes followed by a sixteenth note. In the second measure, there are three eighth notes followed by a sixteenth note. Red double-headed arrows between the staves indicate matching strong beats (SB) between the two strains. Orange double-headed arrows indicate non-matched strong beats (SB) where a strong beat in one strain corresponds to a weak beat in the other. Specifically, the first two eighth notes in both measures have red arrows, while the third eighth note and the sixteenth note have orange arrows.

# What does the algorithm do?

The algorithm compares the first two bars (the “fingerprints”) of two given strains. This comparison is effected through a series of functions that:

- compare strong beats to strong beats and identify **matched SB**
- establish an optimal correspondence between strong beats (=> **non-matched SB**)
- search for **missing strong beats on corresponding weak beats**
- compare weak beats to weak beats and identify **matched WB on matched SB** and **matched WB on non-matched SB**
- compare contour and check for strong-beat reversal on consecutive non-matching strong beats, and then check for alignment between corresponding weak beats

Ex. 2

Allard 1928, A strain

Lajoie 1951, E strain (transposed from C major)

# What does the algorithm do?

The algorithm compares the first two bars (the “fingerprints”) of two given strains. This comparison is effected through a series of functions that:

- compare strong beats to strong beats and identify **matched SB**
- establish an optimal correspondence between strong beats (=> **non-matched SB**)
- search for **missing strong beats on corresponding weak beats**
- compare weak beats to weak beats and identify **matched WB on matched SB** and **matched WB on non-matched SB**
- compare **contour** and check for strong-beat reversal on consecutive non-matching strong beats, and then check for alignment between corresponding weak beats

Ex. 2

Allard 1928, A strain

Lajoie 1951, E strain (transposed from C major)

S7 to S8 = minor 3<sup>rd</sup>

Identical SB contour

# What does the algorithm do?

The algorithm compares the first two bars (the “fingerprints”) of two given strains. This comparison is effected through a series of functions that:

- compare strong beats to strong beats and identify **matched SB**
- establish an optimal correspondence between strong beats ( $\Rightarrow$  **non-matched SB**)
- search for **missing strong beats on corresponding weak beats**
- compare weak beats to weak beats and identify **matched WB on matched SB** and **matched WB on non-matched SB**
- compare **contour** and check for strong-beat reversal on consecutive non-matching strong beats, and then check for **alignment between corresponding weak beats**

Ex. 2

Allard 1928, A strain

Lajoie 1951, E strain (transposed from C major)

S7 to S8 = minor 3<sup>rd</sup>  
S7 to W7 = major 2<sup>nd</sup>

S7 to S8 = minor 3<sup>rd</sup>  
S7 to W7 = minor 2<sup>nd</sup>

Identical SB contour

Similar WB contour

# What does the algorithm do?

The algorithm compares the first two bars (the “fingerprints”) of two given strains. This comparison is effected through a series of functions that:

- compare strong beats to strong beats and identify **matched SB**
- establish an optimal correspondence between strong beats (=> **non-matched SB**)
- search for **missing strong beats on corresponding weak beats**
- compare weak beats to weak beats and identify **matched WB on matched SB** and matched WB on non-matched SB
- compare contour and check for strong-beat reversal on consecutive non-matching strong beats, and then check for alignment between corresponding weak beats

Ex. 3

The image shows two musical staves. The top staff is labeled "Allard 1928, B strain" and the bottom staff is labeled "Joyal 1956, B strain (transposed from A major)". Both staves are in treble clef and have a key signature of one sharp. The notation consists of eighth and sixteenth notes. Colored arrows indicate correspondences between notes in different strains. Red double-headed arrows connect notes in the same position across both staves. Green double-headed arrows connect notes in the same position but with different note values (eighth vs. sixteenth). Orange double-headed arrows connect notes in the same position but with different rhythms (sixteenth vs. eighth). A purple arrow points from a note in the Joyal staff to a note in the Allard staff, indicating a missing strong beat on a corresponding weak beat.

# What does the algorithm do?

The algorithm compares the first two bars (the “fingerprints”) of two given strains. This comparison is effected through a series of functions that:

- compare strong beats to strong beats and identify **matched SB**
- establish an optimal correspondence between strong beats (=> **non-matched SB**)
- search for **missing strong beats on corresponding weak beats**
- compare weak beats to weak beats and identify **matched WB on matched SB** and **matched WB on non-matched SB**
- compare contour and **check for strong-beat reversal on consecutive non-matching strong beats, and then check for alignment between corresponding weak beats**

Ex. 3

Allard 1928, B strain

Joyal 1956, B strain (transposed from A major)

## Correspondences for equal-length fingerprints

If the two fingerprints are the same length, as measured by the number of strong beats, the algorithm establishes a one-to-one correspondence between the strong beats and outputs a list of matched strong beats and a list of non-matched (but corresponding) strong beats. In Example 2, strong beats 1, 2, 3, 5, 6 and 9 are **matched SB** and strong beats 4, 7 and 8 are **corresponding but non-matched SB**.

Given two fingerprints of equal length, all strong beats will be on either the matched or the non-matched list.

Ex. 2

The image shows two staves of musical notation. The top staff is labeled "Allard 1928, A strain" and the bottom staff is labeled "Lajoie 1951, E strain (transposed from C major)". Both staves have a treble clef and a key signature of one sharp. The notation consists of eighth and sixteenth notes. Red double-headed arrows connect strong beats between the two staves. Specifically, red arrows connect the first, second, third, fifth, sixth, and ninth beats of both staves. Orange double-headed arrows connect the fourth, seventh, and eighth beats of both staves, indicating they are corresponding but non-matched strong beats.

## Correspondences for fingerprints of different lengths

If the two fingerprints are different lengths, the algorithm establishes a one-to-one correspondence between as many **matched strong beats** as possible.

Ex. 4

The image shows two musical staves. The top staff is labeled "Boivin 1974, A strain" and the bottom staff is labeled "Soucy 1927, G strain (transposed from A major)". Both staves have a treble clef and a key signature of one sharp. The top staff has a common time signature, while the bottom staff has a 3/2 time signature. Red arrows connect specific notes between the two staves, indicating matched strong beats. One pair of notes is highlighted with red double-headed arrows. The notes are represented by black dots on vertical stems.

Boivin 1974, A strain

Soucy 1927, G strain (transposed from A major)

## Correspondences for fingerprints of different lengths

If the two fingerprints are different lengths, the algorithm establishes a one-to-one correspondence between **as many matched strong beats** as possible.

If the number of non-matched SB between two matched SB is equal in both fingerprints, those **corresponding but non-matched SB** are put on the non-matched strong beat list.

Given two fingerprints of different lengths, some strong beats may not be on either the matched or the non-matched list (i.e., SB3 and SB4 of the Boivin fingerprint below, and SB3, SB4, SB5 and SB6 of the Soucy fingerprint). The weak beat, contour and reversal functions will not apply to non-corresponding strong beats.

Ex. 4

The image shows two musical staves. The top staff is labeled "Boivin 1974, A strain" and the bottom staff is labeled "Soucy 1927, G strain (transposed from A major)". Both staves have a treble clef and a key signature of one sharp. Red arrows indicate correspondences between strong beats. Two red double-headed arrows at the beginning of the Boivin staff point to the first two notes. From the third note of the Boivin staff, three red arrows point to the second, fourth, and fifth notes of the Soucy staff. An orange arrow points from the fifth note of the Boivin staff to the eighth note of the Soucy staff. The Soucy staff has a measure line at the beginning of the third measure.

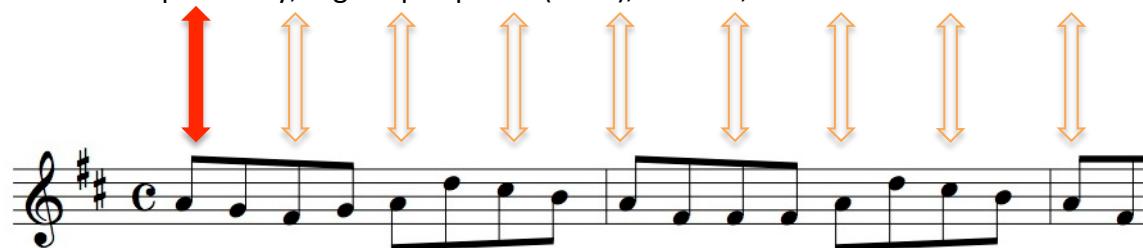
# Truncation

When calculating the above similarity measures, the algorithm first assumes that the **first strong beat matches** (i.e., both fingerprints start on the same pitch).

Ex. 5



Arthur-Joseph Boulay, Gigues pot-pourri (1923), 1<sup>st</sup> tune, A strain



Arthur-Joseph Boulay, Reel à quatre (1929), A strain (transposed to D)

One matched SB  
and eight  
non-matched SB

# Truncation

When calculating the above similarity measures, the algorithm first assumes that the **first strong beat matches** (i.e., both fingerprints start on the same pitch).

It then **truncates the first full beat** (strong beat plus weak beats) of each fingerprint...

Ex. 5

Truncate

The image shows two musical staves. The top staff is from "Gigues pot-pourri" by Arthur-Joseph Boulay, 1923, 1st tune, A strain. The bottom staff is from "Reel à quatre" by Arthur-Joseph Boulay, 1929, A strain (transposed to D). Both staves are in common time with a key signature of one sharp. The first measure of each staff is highlighted with a blue box. A red double-headed arrow on the left side of the top staff indicates the duration of this first measure. A blue box labeled "Truncate" with a downward arrow points to the start of the first measure. Below each staff, orange double-headed arrows indicate the duration of the remaining measures. The top staff has seven orange arrows, and the bottom staff has six.

Arthur-Joseph Boulay, Gigues pot-pourri (1923), 1<sup>st</sup> tune, A strain

Arthur-Joseph Boulay, Reel à quatre (1929), A strain (transposed to D)

# Truncation

When calculating the above similarity measures, the algorithm first assumes that the first strong beat matches (i.e., both fingerprints start on the same pitch).

It then **truncates the first full beat** (strong beat plus weak beats) of each fingerprint, assumes that the **second strong beat matches...**

Ex. 5

Truncate

The image shows two musical staves. The top staff is from "Arthur-Joseph Boulay, Gigues pot-pourri (1923), 1<sup>st</sup> tune, A strain". The bottom staff is from "Arthur-Joseph Boulay, Reel à quatre (1929), A strain". Both staves are in common time (indicated by 'C') and have a key signature of three sharps. A blue vertical bar highlights the first beat of each staff. A red double-headed arrow below the bar indicates the duration of the first beat. A blue box labeled "Truncate" with a blue arrow points to the first beat of the top staff. The rest of the music consists of eighth-note patterns.

# Truncation

When calculating the above similarity measures, the algorithm first assumes that the first strong beat matches (i.e., both fingerprints start on the same pitch).

It then **truncates the first full beat** (strong beat plus weak beats) of each fingerprint, assumes that the **second strong beat matches**, and re-runs the comparison.

Ex. 5

Truncate

The figure shows two musical staves. The top staff is from "Gigues pot-pourri (1923)" by Arthur-Joseph Boulay, and the bottom staff is from "Reel à quatre (1929)" by the same composer. Both staves are in common time (indicated by 'C') and G major (indicated by a treble clef). The first measure of each staff is highlighted with a blue box. Red double-headed arrows point between corresponding notes in the second measure of each staff, indicating a match. A red box on the right contains the text "Eight matched SB".

Arthur-Joseph Boulay, Gigues pot-pourri (1923), 1<sup>st</sup> tune, A strain

Arthur-Joseph Boulay, Reel à quatre (1929), A strain

Eight matched SB

# Truncation

When calculating the above similarity measures, the algorithm first assumes that the first strong beat matches (i.e., both fingerprints start on the same pitch).

It then **truncates the first full beat** (strong beat plus weak beats) of each fingerprint, assumes that the **second strong beat matches**, and re-runs the comparison.

This continues until the number of truncations exceeds 50% of the number of strong beats in the shorter fingerprint. The algorithm outputs all of these comparisons and selects the best result.

Ex. 5

Arthur-Joseph Boulay, Gigues pot-pourri (1923), 1<sup>st</sup> tune, A strain

Arthur-Joseph Boulay, Reel à quatre (1929), A strain

Truncate

Eight matched SB

The figure shows two musical staves. The top staff is from 'Gigues pot-pourri' and the bottom staff is from 'Reel à quatre'. Both staves are in G major (two sharps) and common time. A blue vertical bar labeled 'Truncate' points to the first beat of the top staff. Red double-headed arrows below the staves indicate matching strong beats between the two pieces. There are seven pairs of arrows, with the eighth pair highlighted in red, indicating the point where the algorithm stops truncating because it has matched eight strong beats, which exceeds 50% of the eight strong beats in the shorter staff.

# **Weighting the results**

The algorithm currently measures similarity using the following formula:

**Similarity measure = Number of matching strong beats**  
+ **Number of matching weak beats**  
+ **(Matching contours and reversals)/2**

“Number of matching strong beats” = (number of elements in the “matching strong beats list”)  
+ (number of “strong beats displaced to weak beats”)/2

“Number of matching weak beats” = (number of “matching weak beats on matching strong beats”)\*.75  
+ (number of “matching weak beats on non-matching strong beats”)/2

# Weighting the results

The algorithm currently measures similarity using the following formula:

**Similarity measure = Number of matching strong beats [6 + 1/2]**

$$\begin{aligned} &+ \text{Number of matching weak beats } [5*0.75 + 3*0.5] \\ &+ (\text{Matching contours and reversals } [0])/2 \\ &= 11.75 \end{aligned}$$

“Number of matching strong beats” = (number of elements in the “matching strong beats list”) + (number of “strong beats displaced to weak beats”)/2

“Number of matching weak beats” = (number of “matching weak beats on matching strong beats”)\*.75 + (number of “matching weak beats on non-matching strong beats”)/2

Ex. 1



Allard 1928, A strain

Boivin 1974, A strain

# Weighting the results

The algorithm currently measures similarity using the following formula:

**Similarity measure = Number of matching strong beats [6]**

$$\begin{aligned} &+ \text{Number of matching weak beats } [6 * 0.75 + 2 * 0.5] \\ &+ (\text{Matching contours and reversals } [1 + 0]) / 2 \\ &= 12 \end{aligned}$$

The algorithm currently finds both similar and identical weak-beat contours but only weights the former.

Where:

“Number of matching strong beats” = (number of elements in the “matching strong beats list”) + (number of “strong beats displaced to weak beats”)/2

“Number of matching weak beats” = (number of “matching weak beats on matching strong beats”)\*.75 + (number of “matching weak beats on non-matching strong beats”)/2

Ex. 2

Allard 1928, A strain

Lajoie 1951, E strain (transposed from C major)

S7 to S8 = minor 3<sup>rd</sup>  
S7 to W7 = major 2<sup>nd</sup>

S7 to S8 = minor 3<sup>rd</sup>  
S7 to W7 = minor 2<sup>nd</sup>

Identical SB contour

Similar WB contour

Detailed description: The image shows two musical staves. The top staff is labeled 'Allard 1928, A strain' and the bottom staff is labeled 'Lajoie 1951, E strain (transposed from C major)'. Both staves have a treble clef and a key signature of one sharp. The music consists of eighth notes and sixteenth notes. Colored arrows indicate contours between strong beats (S7 to S8) and weak beats (W7). Red arrows point up from S7 to S8, green arrows point down from S8 to W7, and orange arrows point up from W7 to S8. Some arrows are double-headed. A blue box highlights a segment of the Lajoie staff where a red arrow points up from S7 to S8 and a green arrow points down from S8 to W7. Blue arrows point from this highlighted area to the text 'Identical SB contour' and 'Similar WB contour'. Text annotations above the staves specify 'S7 to S8 = minor 3<sup>rd</sup>' and 'S7 to W7 = major 2<sup>nd</sup>' for Allard, and 'S7 to S8 = minor 3<sup>rd</sup>' and 'S7 to W7 = minor 2<sup>nd</sup>' for Lajoie.

# **How does it work?**

# How does it work?

- The fingerprint
- Matched and non-matched strong beats
- Matched weak beats on matched and non-matched strong beats
- Functions
  - Strong beat displaced to weak beat
  - Matched weak beats on matched strong beats
  - Matched weak beats on non-matched strong beats
  - Identical or similar contour on strong beats
  - Identical or similar contour extends to weak beats
  - Reversal of strong beats
  - Reversal extends to weak beats

# The fingerprint

Given a sample melody:

| **S1 W1 S2 W2 S3 W3 S4 W4 | S5 W5 S6 W6 S7 W7 S8 W8 | S9 W9**

where “S” indicates a note on a strong beat and “W” indicates a note or notes on a weak beat, the “fingerprint” is the numerical array:

	1	2	3	4	5	6	7	8	9
w	S1 => W1	S2 => W2	S3 => W3	S4 => W4	S5 => W5	S6 => W6	S7 => W7	S8 => W8	S9=> W9
0	S1 => S1	S2 => S2	S3 => S3	S4 => S4	S5 => S5	S6 => S6	S7 => S7	S8 => S8	S8 => S9
1	S1 => S2	S2 => S3	S3 => S4	S4 => S5	S5 => S6	S6 => S7	S7 => S8	S8 => S9	
2	S1 => S3	S2 => S4	S3 => S5	S4 => S6	S5 => S7	S6 => S8	S7 => S9		
3	S1 => S4	S2 => S5	S3 => S6	S4 => S7	S5 => S8	S6 => S9			
4	S1 => S5	S2 => S6	S3 => S7	S4 => S8	S5 => S9				
5	S1 => S6	S2 => S7	S3 => S8	S4 => S9					
6	S1 => S7	S2 => S8	S3 => S9						
7	S1 => S8	S2 => S9							
8	S1 => S9								

where “S1=>W1” indicates the interval between S1 and W1, etc.

# The fingerprint

Fingerprint 1:

w	1	2	3	4	5	6	7	8	9
0.0	-3.5	-2.0	-3.5	2.0	-3.5	-2.0	1.0	2.0	-3.5
1.0	0	0	0	0	0	0	0	0	0
2.0	-1.5	1.5	-1	1	-1.5	-1	1.5	1	NaN
2.0	0	0.5	0	-0.5	-2.5	0.5	2.5	NaN	NaN
3.0	-1	1.5	-1.5	-1.5	-1	1.5	NaN	NaN	NaN
4.0	0	0	-2.5	0	0	NaN	NaN	NaN	NaN
5.0	-1.5	-1	-1	1	NaN	NaN	NaN	NaN	NaN
6.0	-2.5	0.5	0	NaN	NaN	NaN	NaN	NaN	NaN
7.0	-1	1.5	NaN	NaN	NaN	NaN	NaN	NaN	NaN
8.0	0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN



Allard 1928, A strain

# The fingerprint

Multiple weak beats

Fingerprint 2:

w	1	2	3	4	5	6	7	8	9
0.0	[-2.5]	[1.0, 1.5]	[2.5]	[5.0]	[-2.5]	[4.5]	[4.0]	[4.5]	[-2.5]
1.0	0	0	0	0	0	0	0	0	0
2.0	-1.5	-1	-1.5	4	-3.5	-0.5	-1	5	NaN
2.0	-2.5	-2.5	2.5	0.5	-4	-1.5	4	NaN	NaN
3.0	-4	1.5	-1	0	-5	3.5	NaN	NaN	NaN
4.0	0	-2	-1.5	-1	0	NaN	NaN	NaN	NaN
5.0	-3.5	-2.5	-2.5	4	NaN	NaN	NaN	NaN	NaN
6.0	-4	-3.5	2.5	NaN	NaN	NaN	NaN	NaN	NaN
7.0	-5	1.5	NaN	NaN	NaN	NaN	NaN	NaN	NaN
8.0	0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN



Allard 1928, B strain

# Matched and non-matched strong beats



Allard 1928, B strain



Joyal 1956, B strain (transposed from A major)

Fingerprint 1:

	1	2	3	4	5	6	7	8	9
w	[-2.5]	[1.0, 1.5]	[2.5]	[5.0]	[-2.5]	[4.5]	[4.0]	[4.5]	[-2.5]
0.0	0	0	0	0	0	0	0	0	0
1.0	-1.5	-1	-1.5	4	-3.5	-0.5	-1	5	NaN
2.0	-2.5	-2.5	2.5	0.5	-4	-1.5	4	NaN	NaN
3.0	-4	1.5	-1	0	-5	3.5	NaN	NaN	NaN
4.0	0	-2	-1.5	-1	0	NaN	NaN	NaN	NaN
5.0	-3.5	-2.5	-2.5	4	NaN	NaN	NaN	NaN	NaN
6.0	-4	-3.5	2.5	NaN	NaN	NaN	NaN	NaN	NaN
7.0	-5	1.5	NaN	NaN	NaN	NaN	NaN	NaN	NaN
8.0	0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Fingerprint 2:

	1	2	3	4	5	6	7	8	9
w	[-2.5]	[1.5]	[0.0]	[1.5]	[-2.5]	[1.5]	[-0.5]	[-1.0]	[-2.5]
0.0	0	0	0	0	0	0	0	0	0
1.0	-1.5	-1	-1.5	4	-4	0.5	-1.5	5	NaN
2.0	-2.5	-2.5	2.5	0	-3.5	-1	3.5	NaN	NaN
3.0	-4	1.5	-1.5	0.5	-5	4	NaN	NaN	NaN
4.0	0	-2.5	-1	-1	0	NaN	NaN	NaN	NaN
5.0	-4	-2	-2.5	4	NaN	NaN	NaN	NaN	NaN
6.0	-3.5	-3.5	2.5	NaN	NaN	NaN	NaN	NaN	NaN
7.0	-5	1.5	NaN	NaN	NaN	NaN	NaN	NaN	NaN
8.0	0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

non-matched  
strong beats

non-matched  
strong beats

# Matched and non-matched strong beats

The image shows two musical staves side-by-side. The left staff, labeled "Allard 1928, A strain", consists of six measures of music in common time with a treble clef. It features eighth-note patterns such as eighth-note pairs, sixteenth-note pairs, and sixteenth-note triplets. The right staff, labeled "Boivin 1974, A strain", also in common time and treble clef, contains six measures of music. It includes eighth-note pairs, sixteenth-note pairs, and sixteenth-note triplets, mirroring the patterns from the Allard version.

non-matched  
strong heats

# Strong beat displaced to weak beat

Question: Has a non-matched strong beat note been displaced to the corresponding weak beat in the other fingerprint?

Solution: Given a non-matched strong beat “SX” in one fingerprint that corresponds to strong beat “SY” in another, add the values “S1 => SX” and “SX => WX” and compare to “S1 => SY”. If they are the same, then weak beat X is the same note as strong beat Y. Note that X=Y if the two fingerprints are of the same length.

Output: “1” for each strong beat displaced to a weak beat. “0” for a non-matched strong beat without displacement. “nan” for matched strong beats.

	1	2	3	4	5	6	7	8	9
w	S1 => W1	S2 => W2	S3 => W3	S4 => W4	S5 => W5	S6 => W6	S7 => W7	S8 => W8	S9 => W9
0	S1 => S1	S2 => S2	S3 => S3	S4 => S4	S5 => S5	S6 => S6	S7 => S7	S8 => S8	S9 => S9
1	S1 => S2	S2 => S3	S3 => S4	S4 => S5	S5 => S6	S6 => S7	S7 => S8	S8 => S9	
2	S1 => S3	S2 => S4	S3 => S5	S4 => S6	S5 => S7	S6 => S8	S7 => S9		
3	S1 => S4	S2 => S5	S3 => S6	S4 => S7	S5 => S8	S6 => S9			
4	S1 => S5	S2 => S6	S3 => S7	S4 => S8	S5 => S9				
5	S1 => S6	S2 => S7	S3 => S8	S4 => S9					
6	S1 => S7	S2 => S8	S3 => S9						
7	S1 => S8	S2 => S9							
8	S1 => S9								

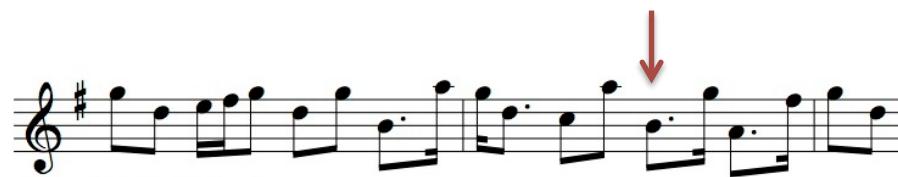
	1	2	3	4	5	6	7	8	9
w	S1 => W1	S2 => W2	S3 => W3	S4 => W4	S5 => W5	S6 => W6	S7 => W7	S8 => W8	S9 => W9
0	S1 => S1	S2 => S2	S3 => S3	S4 => S4	S5 => S5	S6 => S6	S7 => S7	S8 => S8	S9 => S9
1	S1 => S2	S2 => S3	S3 => S4	S4 => S5	S5 => S6	S6 => S7	S7 => S8	S8 => S9	
2	S1 => S3	S2 => S4	S3 => S5	S4 => S6	S5 => S7	S6 => S8	S7 => S9		
3	S1 => S4	S2 => S5	S3 => S6	S4 => S7	S5 => S8	S6 => S9			
4	S1 => S5	S2 => S6	S3 => S7	S4 => S8	S5 => S9				
5	S1 => S6	S2 => S7	S3 => S8	S4 => S9					
6	S1 => S7	S2 => S8	S3 => S9						
7	S1 => S8	S2 => S9							
8	S1 => S9								

non-matched  
strong beat

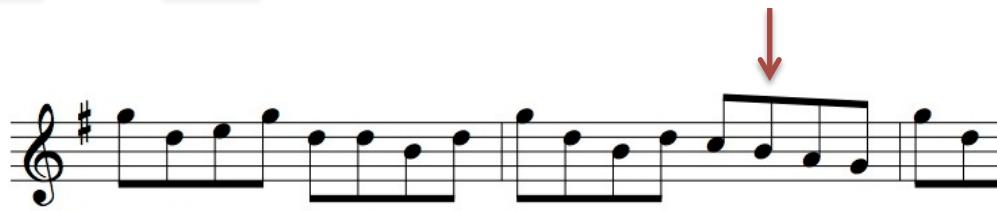
$$\text{[yellow box]} + \text{[yellow box]} = \text{[blue box]} ?$$

# Strong beat displaced to weak beat

 +  = 



Allard 1928, B strain



Joyal 1956, B strain (transposed from A major)

Fingerprint 1:

	1	2	3	4	5	6	7	8	9
w	[-2.5]	[1.0, 1.5]	[2.5]	[5.0]	[-2.5]	[4.5]	[4.0]	[4.5]	[-2.5]
0.0	0	0	0	0	0	0	0	0	0
1.0	-1.5	-1	-1.5	4	-3.5	-0.5	-1	5	NaN
2.0	-2.5	-2.5	2.5	0.5	-4	-1.5	4	NaN	NaN
3.0	-4	1.5	-1	0	-5	3.5	NaN	NaN	NaN
4.0	0	-2	-1.5	-1	0	NaN	NaN	NaN	NaN
5.0	-3.5	-2.5	-2.5	4	NaN	NaN	NaN	NaN	NaN
6.0	-4	-3.5	2.5	NaN	NaN	NaN	NaN	NaN	NaN
7.0	-5	1.5	NaN	NaN	NaN	NaN	NaN	NaN	NaN
8.0	0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Fingerprint 2:

	1	2	3	4	5	6	7	8	9
w	[-2.5]	[1.5]	[0.0]	[1.5]	[-2.5]	[1.5]	[-0.5]	[-1.0]	[-2.5]
0.0	0	0	0	0	0	0	0	0	0
1.0	-1.5	-1	-1.5	4	-4	0.5	-1.5	5	NaN
2.0	-2.5	-2.5	2.5	0	-3.5	-1	3.5	NaN	NaN
3.0	-4	1.5	-1.5	0.5	-5	4	NaN	NaN	NaN
4.0	0	-2.5	-1	-1	0	NaN	NaN	NaN	NaN
5.0	-4	-2	-2.5	4	NaN	NaN	NaN	NaN	NaN
6.0	-3.5	-3.5	2.5	NaN	NaN	NaN	NaN	NaN	NaN
7.0	-5	1.5	NaN	NaN	NaN	NaN	NaN	NaN	NaN
8.0	0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

non-matched  
strong beat



# Matched weak beats on matched strong beats

Question: Given a matched strong beat, do the corresponding weak beats also match?

Solution: Given a matched strong beat “SX” in one fingerprint that corresponds to strong beat “SY” in another, compare the elements of “SX => WX” with those of “SY => WY”.

If WX and/or WY contain multiple elements, look for 1) the number of elements in common and 2) whether or not those common elements are in the same order in both. Each of these outputs is worth 50% of the total.

Maximum output per strong beat: 1 (“nan” = non-matched strong beat)

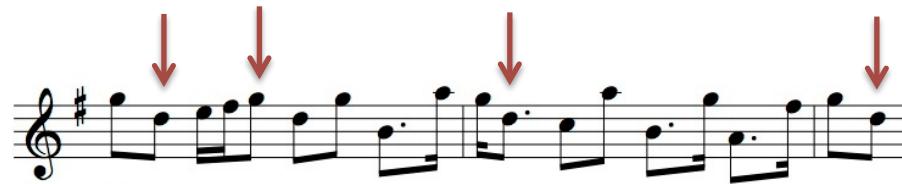
w	1	2	3	4	5	6	7	8	9
0	S1 => W1	S2 => W2	S3 => W3	S4 => W4	S5 => W5	S6 => W6	S7 => W7	S8 => W8	S9 => W9
1	S1 => S1	S2 => S2	S3 => S3	S4 => S4	S5 => S5	S6 => S6	S7 => S7	S8 => S8	S9 => S9
2	S1 => S2	S2 => S3	S3 => S4	S4 => S5	S5 => S6	S6 => S7	S7 => S8	S8 => S9	
3	S1 => S3	S2 => S4	S3 => S5	S4 => S6	S5 => S7	S6 => S8	S7 => S9		
4	S1 => S4	S2 => S5	S3 => S6	S4 => S7	S5 => S8	S6 => S9			
5	S1 => S5	S2 => S6	S3 => S7	S4 => S8	S5 => S9				
6	S1 => S6	S2 => S7	S3 => S8	S4 => S9					
7	S1 => S7	S2 => S8	S3 => S9						
8	S1 => S8	S2 => S9							

w	1	2	3	4	5	6	7	8	9
0	S1 => W1	S2 => W2	S3 => W3	S4 => W4	S5 => W5	S6 => W6	S7 => W7	S8 => W8	S9 => W9
1	S1 => S1	S2 => S2	S3 => S3	S4 => S4	S5 => S5	S6 => S6	S7 => S7	S8 => S8	S9 => S9
2	S1 => S2	S2 => S3	S3 => S4	S4 => S5	S5 => S6	S6 => S7	S7 => S8	S8 => S9	
3	S1 => S3	S2 => S4	S3 => S5	S4 => S6	S5 => S7	S6 => S8	S7 => S9		
4	S1 => S4	S2 => S5	S3 => S6	S4 => S7	S5 => S8	S6 => S9			
5	S1 => S5	S2 => S6	S3 => S7	S4 => S8	S5 => S9				
6	S1 => S6	S2 => S7	S3 => S8	S4 => S9					
7	S1 => S7	S2 => S8	S3 => S9						
8	S1 => S8	S2 => S9							

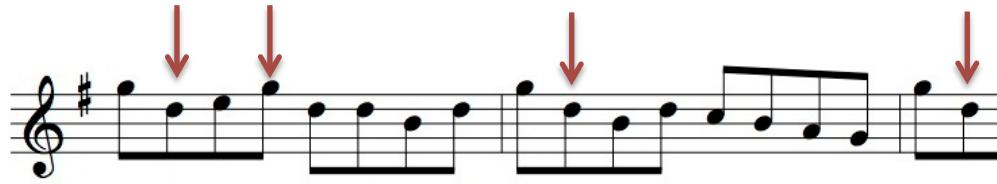
matched  
strong beat

= ?

# Matched weak beats on matched strong beats



Allard 1928, B strain



Joyal 1956, B strain (transposed from A major)

Fingerprint 1:

	1	2	3	4	5	6	7	8	9
w	[-2.5]	[1.0, 1.5]	[2.5]	[5.0]	[-2.5]	[4.5]	[4.0]	[4.5]	[-2.5]
0.0	0	0	0	0	0	0	0	0	0
1.0	-1.5	-1	-1.5	4	-3.5	-0.5	-1	5	NaN
2.0	-2.5	-2.5	2.5	0.5	-4	-1.5	4	NaN	NaN
3.0	-4	1.5	-1	0	-5	3.5	NaN	NaN	NaN
4.0	0	-2	-1.5	-1	0	NaN	NaN	NaN	NaN
5.0	-3.5	-2.5	-2.5	4	NaN	NaN	NaN	NaN	NaN
6.0	-4	-3.5	2.5	NaN	NaN	NaN	NaN	NaN	NaN
7.0	-5	1.5	NaN	NaN	NaN	NaN	NaN	NaN	NaN
8.0	0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Fingerprint 2:

	1	2	3	4	5	6	7	8	9
w	[-2.5]	[1.5]	[0.0]	[1.5]	[-2.5]	[1.5]	[-0.5]	[-1.0]	[-2.5]
0.0	0	0	0	0	0	0	0	0	0
1.0	-1.5	-1	-1.5	4	-4	0.5	-1.5	5	NaN
2.0	-2.5	-2.5	2.5	0	-3.5	-1	3.5	NaN	NaN
3.0	-4	1.5	-1.5	0.5	-5	4	NaN	NaN	NaN
4.0	0	-2.5	-1	-1	0	NaN	NaN	NaN	NaN
5.0	-4	-2	-2.5	4	NaN	NaN	NaN	NaN	NaN
6.0	-3.5	-3.5	2.5	NaN	NaN	NaN	NaN	NaN	NaN
7.0	-5	1.5	NaN	NaN	NaN	NaN	NaN	NaN	NaN
8.0	0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

non-matched  
strong beats



# Matched weak beats on non-matched strong beats

Question: Given a non-matched strong beat, do the corresponding weak beats match?

Solution: Given a non-matched strong beat “SX” in one fingerprint that corresponds to strong beat “SY” in another, add “S1 => SX” and “SX => WX”. Compare that result to “S1 => SY” plus “SY => WY”. If they are the same, then weak beat X is the same note as weak beat Y.

If WX and/or WY contain multiple elements, look for 1) the number of elements in common and 2) whether or not those common elements are in the same order in both. Each of these outputs is worth 50% of the total.

Maximum output per strong beat: 1 (“nan” = matching strong beat)

w	1	2	3	4	5	6	7	8	9
0	S1 => W1	S2 => W2	S3 => W3	S4 => W4	S5 => W5	S6 => W6	S7 => W7	S8 => W8	S9=> W9
1	S1 => S1	S2 => S2	S3 => S3	S4 => S4	S5 => S5	S6 => S6	S7 => S7	S8 => S8	S8 => S9
2	S1 => S2	S2 => S3	S3 => S4	S4 => S5	S5 => S6	S6 => S7	S7 => S8	S8 => S9	
3	S1 => S3	S2 => S4	S3 => S5	S4 => S6	S5 => S7	S6 => S8	S7 => S9		
4	S1 => S4	S2 => S5	S3 => S6	S4 => S7	S5 => S8	S6 => S9			
5	S1 => S5	S2 => S6	S3 => S7	S4 => S8	S5 => S9				
6	S1 => S6	S2 => S7	S3 => S8	S4 => S9					
7	S1 => S7	S2 => S8	S3 => S9						
8	S1 => S8	S2 => S9							
	S1 => S9								

w	1	2	3	4	5	6	7	8	9
0	S1 => W1	S2 => W2	S3 => W3	S4 => W4	S5 => W5	S6 => W6	S7 => W7	S8 => W8	S9=> W9
1	S1 => S1	S2 => S2	S3 => S3	S4 => S4	S5 => S5	S6 => S6	S7 => S7	S8 => S8	S8 => S9
2	S1 => S2	S2 => S3	S3 => S4	S4 => S5	S5 => S6	S6 => S7	S7 => S8	S8 => S9	
3	S1 => S3	S2 => S4	S3 => S5	S4 => S6	S5 => S7	S6 => S8	S7 => S9		
4	S1 => S4	S2 => S5	S3 => S6	S4 => S7	S5 => S8	S6 => S9			
5	S1 => S5	S2 => S6	S3 => S7	S4 => S8	S5 => S9				
6	S1 => S6	S2 => S7	S3 => S8	S4 => S9					
7	S1 => S7	S2 => S8	S3 => S9						
8	S1 => S8	S2 => S9							
	S1 => S9								

non-matched  
strong beat

$$+ \quad + \quad = \quad + \quad ?$$

8

## **Matched weak beats on non-matched strong beats**

 +  =  + 

## Fingerprint 1:

non-matched  
strong heats

## Fingerprint 2:

# Identical or similar contour on strong beats

Question: Given two consecutive non-matched strong beats, check for identical or similar contour.

Solution: Given non-matched strong beats “SX” and “SX+1” corresponding to “SY” and “SY+1”, compare “SX => SX+1” to “SY => SY+1”. If they have the same value, the contour is identical. If the values are either +.5 or -.5 apart, there is a similar contour. (The half-step difference allows for the figure to start on a different scale degree.) The algorithm also compares consecutive similar contours to determine if they are similar in the same direction (i.e. +.5 or -.5).

Output: 1 for an identical contour, .5 for a similar contour, 0 for consecutive non-matched strong beats with no matched contour, “nan” for everything else (matched strong beats or non-consecutive non-matched strong beats)

	1	2	3	4	5	6	7	8	9
w	S1 => W1	S2 => W2	S3 => W3	S4 => W4	S5 => W5	S6 => W6	S7 => W7	S8 => W8	S9 => W9
0	S1 => S1	S2 => S2	S3 => S3	S4 => S4	S5 => S5	S6 => S6	S7 => S7	S8 => S8	S9 => S9
1	S1 => S2	S2 => S3	S3 => S4	S4 => S5	S5 => S6	S6 => S7	S7 => S8	S8 => S9	
2	S1 => S3	S2 => S4	S3 => S5	S4 => S6	S5 => S7	S6 => S8	S7 => S9		
3	S1 => S4	S2 => S5	S3 => S6	S4 => S7	S5 => S8	S6 => S9			
4	S1 => S5	S2 => S6	S3 => S7	S4 => S8	S5 => S9				
5	S1 => S6	S2 => S7	S3 => S8	S4 => S9					
6	S1 => S7	S2 => S8	S3 => S9						
7	S1 => S8	S2 => S9							
8	S1 => S9								

consecutive  
non-matched  
strong beats

	1	2	3	4	5	6	7	8	9
w	S1 => W1	S2 => W2	S3 => W3	S4 => W4	S5 => W5	S6 => W6	S7 => W7	S8 => W8	S9 => W9
0	S1 => S1	S2 => S2	S3 => S3	S4 => S4	S5 => S5	S6 => S6	S7 => S7	S8 => S8	S9 => S9
1	S1 => S2	S2 => S3	S3 => S4	S4 => S5	S5 => S6	S6 => S7	S7 => S8	S8 => S9	
2	S1 => S3	S2 => S4	S3 => S5	S4 => S6	S5 => S7	S6 => S8	S7 => S9		
3	S1 => S4	S2 => S5	S3 => S6	S4 => S7	S5 => S8	S6 => S9			
4	S1 => S5	S2 => S6	S3 => S7	S4 => S8	S5 => S9				
5	S1 => S6	S2 => S7	S3 => S8	S4 => S9					
6	S1 => S7	S2 => S8	S3 => S9						
7	S1 => S8	S2 => S9							
8	S1 => S9								

$$= \quad +/- \quad 0.5 \quad ?$$

**Identical or similar contour extends to weak beats**

Question: Given an identical or similar contour, as above, do the corresponding weak beats also have an identical or similar contour?

Solution: If SX and SY each only have 1 weak beat, compare “SBX => WBX” to “SBY => WBY”. If they are identical, the weak beat contour is identical. If the values are either +.5 or -.5 apart, there is a similar contour.

Note: If SX and/or SY has multiple weak beats, the algorithm will check for identical contour (1 or more of the same weak beats, appearing in the same order). It does not check for similar contour in this case.

Output: 1 for an identical contour [.5 for a similar contour].

	1	2	3	4	5	6	7	8	9
w	S1 => W1	S2 => W2	S3 => W3	S4 => W4	S5 => W5	S6 => W6	S7 => W7	S8 => W8	S9 => W9
0	S1 => S1	S2 => S2	S3 => S3	S4 => S4	S5 => S5	S6 => S6	S7 => S7	S8 => S8	S9 => S9
1	S1 => S2	S2 => S3	S3 => S4	S4 => S5	S5 => S6	S6 => S7	S7 => S8	S8 => S9	
2	S1 => S3	S2 => S4	S3 => S5	S4 => S6	S5 => S7	S6 => S8	S7 => S9		
3	S1 => S4	S2 => S5	S3 => S6	S4 => S7	S5 => S8	S6 => S9			
4	S1 => S5	S2 => S6	S3 => S7	S4 => S8	S5 => S9				
5	S1 => S6	S2 => S7	S3 => S8	S4 => S9					
6	S1 => S7	S2 => S8	S3 => S9						
7	S1 => S8	S2 => S9							
8	S1 => S9								

consecutive

consecutive  
non-matched  
strong beats  
with identical or  
similar contour

=      +/- 0.5 ?

# Reversal of strong beats

Question: Given two consecutive non-matched strong beats, are they in fact matched strong beats that have been reversed?

Solution: Given non-matched strong beats “SX” and “SX+1” corresponding to “SY” and “SY+1”, does “S1 => SX” equal “S1 => SY+1” and “S1 => SY” equal “S1 => SX+1”? If both are true, then the strong beats are reversed.

Output: “1” if reversed, “0” if non-matched and not reversed, “nan” for everything else (matched strong beats or non-consecutive non-matched strong beats)

	1	2	3	4	5	6	7	8	9
w	S1 => W1	S2 => W2	S3 => W3	S4 => W4	S5 => W5	S6 => W6	S7 => W7	S8 => W8	S9 => W9
0	S1 => S1	S2 => S2	S3 => S3	S4 => S4	S5 => S5	S6 => S6	S7 => S7	S8 => S8	S9 => S9
1	S1 => S2	S2 => S3	S3 => S4	S4 => S5	S5 => S6	S6 => S7	S7 => S8	S8 => S9	
2	S1 => S3	S2 => S4	S3 => S5	S4 => S6	S5 => S7	S6 => S8	S7 => S9		
3	S1 => S4	S2 => S5	S3 => S6	S4 => S7	S5 => S8	S6 => S9			
4	S1 => S5	S2 => S6	S3 => S7	S4 => S8	S5 => S9				
5	S1 => S6	S2 => S7	S3 => S8	S4 => S9					
6	S1 => S7	S2 => S8	S3 => S9						
7	S1 => S8	S2 => S9							
8	S1 => S9								

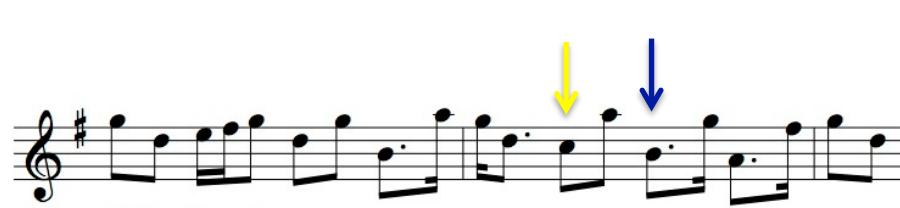
consecutive  
non-matched  
strong beats

	1	2	3	4	5	6	7	8	9
w	S1 => W1	S2 => W2	S3 => W3	S4 => W4	S5 => W5	S6 => W6	S7 => W7	S8 => W8	S9 => W9
0	S1 => S1	S2 => S2	S3 => S3	S4 => S4	S5 => S5	S6 => S6	S7 => S7	S8 => S8	S9 => S9
1	S1 => S2	S2 => S3	S3 => S4	S4 => S5	S5 => S6	S6 => S7	S7 => S8	S8 => S9	
2	S1 => S3	S2 => S4	S3 => S5	S4 => S6	S5 => S7	S6 => S8	S7 => S9		
3	S1 => S4	S2 => S5	S3 => S6	S4 => S7	S5 => S8	S6 => S9			
4	S1 => S5	S2 => S6	S3 => S7	S4 => S8	S5 => S9				
5	S1 => S6	S2 => S7	S3 => S8	S4 => S9					
6	S1 => S7	S2 => S8	S3 => S9						
7	S1 => S8	S2 => S9							
8	S1 => S9								

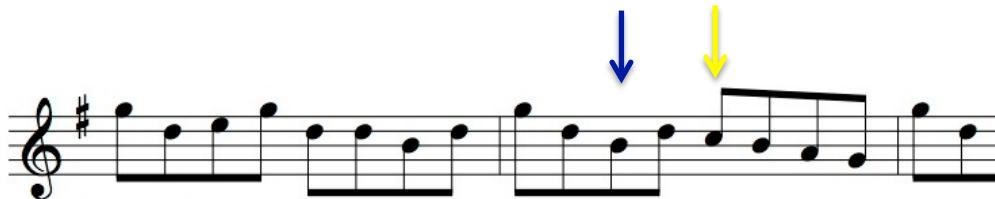
=      and      =      ?

# Reversal of strong beats

= and =



Allard 1928, B strain



Joyal 1956, B strain (transposed from A major)

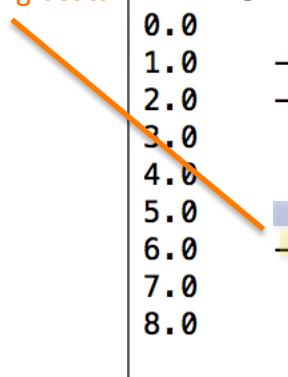
Fingerprint 1:

	1	2	3	4	5	6	7	8	9
w	[-2.5]	[1.0, 1.5]	[2.5]	[5.0]	[-2.5]	[4.5]	[4.0]	[4.5]	[-2.5]
0.0	0	0	0	0	0	0	0	0	0
1.0	-1.5	-1	-1.5	4	-3.5	-0.5	-1	5	NaN
2.0	-2.5	-2.5	2.5	0.5	-4	-1.5	4	NaN	NaN
3.0	-4	1.5	-1	0	-5	3.5	NaN	NaN	NaN
4.0	0	-2	-1.5	-1	0	NaN	NaN	NaN	NaN
5.0	-3.5	-2.5	-2.5	4	NaN	NaN	NaN	NaN	NaN
6.0	-4	-3.5	2.5	NaN	NaN	NaN	NaN	NaN	NaN
7.0	-5	1.5	NaN	NaN	NaN	NaN	NaN	NaN	NaN
8.0	0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Fingerprint 2:

	1	2	3	4	5	6	7	8	9
w	[-2.5]	[1.5]	[0.0]	[1.5]	[-2.5]	[1.5]	[-0.5]	[-1.0]	[-2.5]
0.0	0	0	0	0	0	0	0	0	0
1.0	-1.5	-1	-1.5	4	-4	0.5	-1.5	5	NaN
2.0	-2.5	-2.5	2.5	0	-3.5	-1	3.5	NaN	NaN
3.0	-4	1.5	-1.5	0.5	-5	4	NaN	NaN	NaN
4.0	0	-2.5	-1	-1	0	NaN	NaN	NaN	NaN
5.0	-4	-2	-2.5	4	NaN	NaN	NaN	NaN	NaN
6.0	-3.5	-3.5	2.5	NaN	NaN	NaN	NaN	NaN	NaN
7.0	-5	1.5	NaN	NaN	NaN	NaN	NaN	NaN	NaN
8.0	0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

consecutive  
non-matched  
strong beats



## **Reversal extends to weak beats**

Question: Given two consecutive non-matched strong beats that are reversed (as above), are the weak beats also reversed?

Solution: Does “ $SX \Rightarrow WX$ ” equal “ $SY+1 \Rightarrow WY+1$ ”? If so, one of the weak beats is reversed. Does “ $SX+1 \Rightarrow WX+1$ ” equal “ $SY \Rightarrow WY$ ”? If so, the other weak beat is reversed.

If WX and/or WY contain multiple elements, look for 1) the number of elements in common and 2) whether or not those common elements are in the same order in both. Each of these outputs is worth 50% of the total.

Maximum output per strong beat: 1

	1	2	3	4	5	6	7	8	9
w	S1 => W1	S2 => W2	S3 => W3	S4 => W4	S5 => W5	S6 => W6	S7 => W7	S8 => W8	S9 => W9
0	S1 => S1	S2 => S2	S3 => S3	S4 => S4	S5 => S5	S6 => S6	S7 => S7	S8 => S8	S9 => S9
1	S1 => S2	S2 => S3	S3 => S4	S4 => S5	S5 => S6	S6 => S7	S7 => S8	S8 => S9	
2	S1 => S3	S2 => S4	S3 => S5	S4 => S6	S5 => S7	S6 => S8	S7 => S9		
3	S1 => S4	S2 => S5	S3 => S6	S4 => S7	S5 => S8	S6 => S9			
4	S1 => S5	S2 => S6	S3 => S7	S4 => S8	S5 => S9				
5	S1 => S6	S2 => S7	S3 => S8	S4 => S9					
6	S1 => S7	S2 => S8	S3 => S9						
7	S1 => S8	S2 => S9							
8	S1 => S9								

consecutive

consecutive  
non-matched  
strong beats that  
are reversed

$$\textcolor{yellow}{\boxed{\quad}} = \textcolor{green}{\boxed{\quad}} ?$$

 =  ?

# **Results**

# Results

	1	2	3	4	5	6	7	8	9
Strong Beat Comparison	[0.0, 0, 0]	[-1.5, 1, 1]	[nan, 2, 2]	[nan, 3, 3]	[0.0, 4, 4]	[-1.5, 5, 5]	[-2.5, 6, 6]	[nan, 7, 7]	[0.0, 8, 8]
Displacement Comparison (Strong-Weak)	NaN	NaN	1	0	NaN	NaN	NaN	0	NaN
Weak Beats Comparison (Matched Strong)	[1.0, 1.0]	[1.0, 1.0]	NaN	NaN	[1.0, 1.0]	[1.0, 1.0]	[0, 0]	NaN	[1.0, 1.0]
Weak Beats Comparison (Mismatched Strong)	NaN	NaN	[1.0, 1.0]	[1.0, 1.0]	NaN	NaN	NaN	[1.0, 1.0]	NaN
Contour Comparison (Strong)	NaN	NaN	0	NaN	NaN	NaN	NaN	NaN	NaN
Contour Comparison (Weak)	NaN	NaN	0	NaN	NaN	NaN	NaN	NaN	NaN
Reversal Comparison (Strong)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Reversal Comparison (Weak)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Number of Truncations	0	0	0	0	0	0	0	0	0

Ex. 1

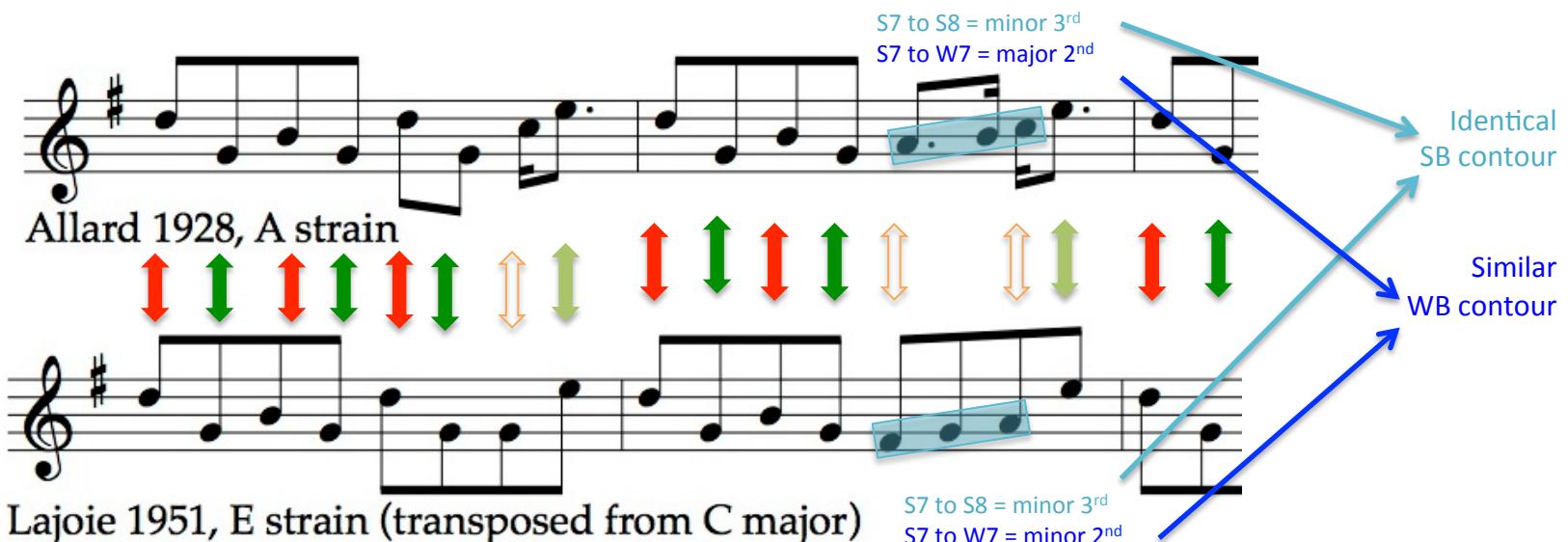
Allard 1928, A strain

Boivin 1974, A strain

# Results

	1	2	3	4	5	6	7	8	9
Strong Beat Comparison	[0.0, 0, 0]	[-1.5, 1, 1]	[0.0, 2, 2]	[nan, 3, 3]	[0.0, 4, 4]	[-1.5, 5, 5]	[nan, 6, 6]	[nan, 7, 7]	[0.0, 8, 8]
Displacement Comparison (Strong-Weak)	NaN	NaN	NaN	0	NaN	NaN	0	0	NaN
Weak Beats Comparison (Matched Strong)	[1.0, 1.0]	[1.0, 1.0]	[1.0, 1.0]	NaN	[1.0, 1.0]	[1.0, 1.0]	NaN	NaN	[1.0, 1.0]
Weak Beats Comparison (Mismatched Strong)	NaN	NaN	NaN	[1.0, 1.0]	NaN	NaN	[0, 0]	[1.0, 1.0]	NaN
Contour Comparison (Strong)	NaN	NaN	NaN	NaN	NaN	NaN	1	NaN	NaN
Contour Comparison (Weak)	NaN	NaN	NaN	NaN	NaN	NaN	0	NaN	NaN
Reversal Comparison (Strong)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Reversal Comparison (Weak)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Number of Truncations	0	0	0	0	0	0	0	0	0

Ex. 2



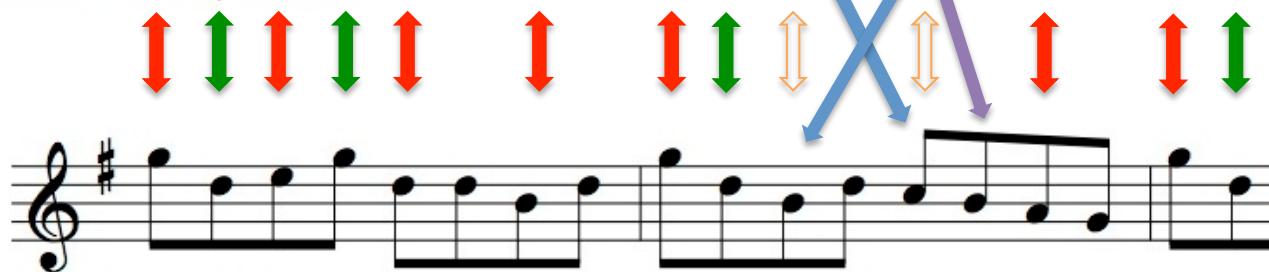
# Results

	Best Result:									
	1	2	3	4	5	6	7	8	9	
Strong Beat Comparison	[0.0, 0, 0]	[-1.5, 1, 1]	[-2.5, 2, 2]	[-4.0, 3, 3]	[0.0, 4, 4]	[nan, 5, 5]	[nan, 6, 6]	[-5.0, 7, 7]	[0.0, 8, 8]	
Displacement Comparison (Strong-Weak)	NaN	NaN	NaN	NaN	NaN	0	1	NaN	NaN	
Weak Beats Comparison (Matched Strong)	[1.0, 1.0]	[1.0, 1.0]	[0, 0]	[0, 0]	[1.0, 1.0]	Nan	NaN	[0, 0]	[1.0, 1.0]	
Weak Beats Comparison (Mismatched Strong)	NaN	NaN	NaN	NaN	NaN	[0, 0]	[0, 0]	NaN	NaN	
Contour Comparison (Strong)	NaN	NaN	NaN	NaN	NaN	0	NaN	NaN	NaN	
Contour Comparison (Weak)	NaN	NaN	NaN	NaN	NaN	0	NaN	NaN	NaN	
Reversal Comparison (Strong)	NaN	NaN	NaN	NaN	NaN	1	NaN	NaN	NaN	
Reversal Comparison (Weak)	NaN	NaN	NaN	NaN	NaN	[[0, 0], [0, 0]]	NaN	NaN	NaN	
Number of Truncations	0	0	0	0	0	0	0	0	0	0

Ex. 3



Allard 1928, B strain



Joyal 1956, B strain (transposed from A major)

# Results

	Best Result:					
	1	2	3	4	5	6
Strong Beat Comparison	[0.0, 0, 0]	[-1.5, 1, 1]	[0.0, 4, 6]	[-1.5, 5, 7]	[-2.5, 6, 8]	[0.0, 8, 10]
Displacement Comparison (Strong-Weak)	NaN	NaN	NaN	NaN	NaN	NaN
Weak Beats Comparison (Matched Strong)	[1.0, 1.0]	[1.0, 1.0]	[1.0, 1.0]	[1.0, 1.0]	[0, 0]	[1.0, 1.0]
Weak Beats Comparison (Mismatched Strong)	NaN	NaN	NaN	NaN	NaN	NaN
Contour Comparison (Strong)	NaN	NaN	NaN	NaN	NaN	NaN
Contour Comparison (Weak)	NaN	NaN	NaN	NaN	NaN	NaN
Reversal Comparison (Strong)	NaN	NaN	NaN	NaN	NaN	NaN
Reversal Comparison (Weak)	NaN	NaN	NaN	NaN	NaN	NaN
Number of Truncations	0	0	0	0	0	0

Ex. 4

The image shows two musical staves. The top staff is labeled "Boivin 1974, A strain" and the bottom staff is labeled "Soucy 1927, G strain (transposed from A major)". Both staves are in 3/2 time with a key signature of one sharp. The notes are represented by vertical stems pointing down. Colored arrows indicate specific note correspondences: red double-headed arrows at the beginning of each staff point to the first two notes; green double-headed arrows point to the next two notes; a red arrow points to the third note; a green arrow points to the fourth note; an orange arrow points to the fifth note; a green arrow points to the sixth note; and a red arrow points to the seventh note. A blue arrow points from the eighth note of the Soucy staff up to a blue box containing the text "\*not in results".

\*not in results

# Results

	Best Result:								
	1	2	3	4	5	6	7	8	
Strong Beat Comparison	[0.0, 0, 0]	[1.5, 1, 1]	[3.5, 2, 2]	[1.5, 3, 3]	[0.0, 4, 4]	[1.5, 5, 5]	[3.5, 6, 6]	[1.5, 7, 7]	
Displacement Comparison (Strong-Weak)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
Weak Beats Comparison (Matched Strong)	[1.0, 1.0]	[1.0, 1.0]	[1.0, 1.0]	[1.0, 1.0]	[1.0, 1.0]	[1.0, 1.0]	[1.0, 1.0]	[1.0, 1.0]	
Weak Beats Comparison (Mismatched Strong)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
Contour Comparison (Strong)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
Contour Comparison (Weak)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
Reversal Comparison (Strong)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
Reversal Comparison (Weak)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
Number of Truncations	1	1	1	1	1	1	1	1	

Ex. 5

Truncate

Arthur-Joseph Boulay, Gigues pot-pourri (1923), 1<sup>st</sup> tune, A strain

Arthur-Joseph Boulay, Reel à quatre (1929), A strain

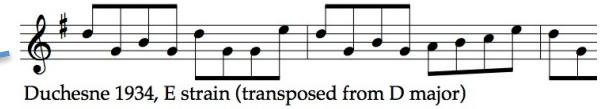
# **Rankings**

# Rankings

```
>>> db.rankings_for('Allard_1928_MoneyMusk_A.xml', 40)
Soucy_1927_MoneyMusk_G.xml           15.75
Soucy_1927_MoneyMusk_O.xml           15.75
Soucy_1927_MoneyMusk_C.xml           15.75
Soucy_1927_MoneyMusk_A.xml           15.75
Boulay_1923_GiguesPotPourri_2ndTune_A.xml 15.75
Soucy_1927_MoneyMusk_K.xml           15.75
Soucy_1927_MoneyMusk_N.xml           15.75
Soucy_1927_MoneyMusk_M.xml           15.75
Picard_1930_MoneyMusk_C.xml          15
Duchesne_1938_MoneyMuskAmericain_A.xml 15
Picard_1930_MoneyMusk_E.xml          15
Soucy_1927_MoneyMusk_E.xml           14.5
Duchesne_1934_MoneyMusk_A.xml          14.5
Duchesne_1934_MoneyMusk_E.xml          14.5
Soucy_1927_MoneyMusk_I.xml           14.5
Potvin_198?_MoneyMusk_J.xml          14.25
Joyal_1956_MoneyMusk_F.xml           14.25
Joyal_1956_MoneyMusk_A.xml           14.25
Potvin_198?_MoneyMusk_L.xml          14.25
Picard_1930_MoneyMusk_A.xml          14.25
Boivin_YEAR_MoneyMusk_F.xml          14
Montmarquette_1928_MoneyMusk_A.xml    13.5
Montmarquette_1928_MoneyMusk_C.xml    13.25
Lajoie_YEAR_MoneyMusk_C.xml          13
Potvin_198?_MoneyMusk_H.xml          12.75
Potvin_198?_MoneyMusk_F.xml          12.75
Potvin_198?_MoneyMusk_B.xml          12.75
Lajoie_YEAR_MoneyMusk_E.xml          12
Boivin_YEAR_MoneyMusk_A.xml          11.75
Joyal_1956_MoneyMusk_H.xml           11.5
Joyal_1956_MoneyMusk_D.xml           11.5
Lajoie_YEAR_MoneyMusk_A.xml          11.5
Richard_1975_MoneyMusk_A.xml          11.5
Boivin_YEAR_MoneyMusk_C.xml          11.25
Potvin_198?_MoneyMusk_D.xml          10.75
Duchesne_1934_MoneyMusk_D.xml          10.5
Soucy_1925_GiguesIrlandaisesNo1Tune1_B.xml 10.25
Soucy_1929_GigueATiGus_B.xml          9.75
Allard_1928_ReelDuVoyageur_B.xml      9.25
Allard_1932_ReelTartesAuxPommes_B.xml 9.25
Name: Allard_1928_MoneyMusk_A.xml, dtype: object
```

# Rankings

```
>>> db.rankings_for('Allard_1928_MoneyMusk_A.xml', 40)
Soucy_1927_MoneyMusk_G.xml
Soucy_1927_MoneyMusk_O.xml
Soucy_1927_MoneyMusk_C.xml
Soucy_1927_MoneyMusk_A.xml
Boulay_1923_GiguesPotPourri_2ndTune_A.xml
Soucy_1927_MoneyMusk_K.xml
Soucy_1927_MoneyMusk_N.xml
Soucy_1927_MoneyMusk_M.xml
Picard_1930_MoneyMusk_C.xml
Duchesne_1938_MoneyMuskAmericain_A.xml
Picard_1930_MoneyMusk_E.xml
Soucy_1927_MoneyMusk_E.xml
Duchesne_1934_MoneyMusk_A.xml
Duchesne_1934_MoneyMusk_E.xml
Soucy_1927_MoneyMusk_I.xml
Potvin_198?_MoneyMusk_J.xml
Joyal_1956_MoneyMusk_F.xml
Joyal_1956_MoneyMusk_A.xml
Potvin_198?_MoneyMusk_L.xml
Picard_1930_MoneyMusk_A.xml
Boivin_YEAR_MoneyMusk_F.xml
Montmarquette_1928_MoneyMusk_A.xml
Montmarquette_1928_MoneyMusk_C.xml
Lajoie_YEAR_MoneyMusk_C.xml
Potvin_198?_MoneyMusk_H.xml
Potvin_198?_MoneyMusk_F.xml
Potvin_198?_MoneyMusk_B.xml
Lajoie_YEAR_MoneyMusk_E.xml
Boivin_YEAR_MoneyMusk_A.xml
Joyal_1956_MoneyMusk_H.xml
Joyal_1956_MoneyMusk_D.xml
Lajoie_YEAR_MoneyMusk_A.xml
Richard_1975_MoneyMusk_A.xml
Boivin_YEAR_MoneyMusk_C.xml
Potvin_198?_MoneyMusk_D.xml
Duchesne_1934_MoneyMusk_D.xml
Soucy_1925_GiguesIrlandaisesNo1Tune1_B.xml
Soucy_1929_GigueATiGus_B.xml
Allard_1928_ReelDuVoyageur_B.xml
Allard_1932_ReelTartesAuxPommes_B.xml
Name: Allard_1928_MoneyMusk_A.xml, dtype: object
```



# Rankings

>>> db.rankings_for('Boulay_1923_GiguesPotPourri_1stTune_A.xml')	
Boulay_1929_ReelAQuatre_A.xml	14
Boulay_1929_ReelAQuatre_B.xml	8
Boulay_1923_GiguesPotPourri_1stTune_B.xml	8
Boulay_1929_ReelDuLabourer_A.xml	7.5
Allard_1929_ReelDuGrandpere_A.xml	7.5
Boivin_YEAR_MoneyMusk_C.xml	7.25
Potvin_198?_MoneyMusk_D.xml	7.25
Boivin_YEAR_MoneyMusk_F.xml	7
Soucy_1929_SetDuCanada3ePartie_A.xml	6.75
Allard_1928_ReelDeLaVeuve_A.xml	6.75
Joyal_1956_MoneyMusk_A.xml	6.25
Lajoie_YEAR_MoneyMusk_E.xml	6.25
Boulay_1928_QuadrilleDeBerthier2ePartie_B.xml	6.25
Picard_1930_MoneyMusk_C.xml	6.25
Montmarquette_1928_MoneyMusk_A.xml	6.25
...	
SoucyLafleur_1929_MarcheDesMarguilliers_B.xml	1.5
Allard_1929_ReelDuSemeur_B.xml	1.5
Soucy_1931_QuadrilleDesAncetres4eFigure_A.xml	1.5
Soucy_1928_QuadrilleNational2eFigure_B.xml	1.5
Boulay_1924_QuadrilleCanadien2ePartie_B.xml	1.5
Allard_1936_QuadrilleFrancais_A.xml	1.25
Allard_1929_QuadrilleDeBeauharnois_A.xml	1
Boulay_1924_QuadrilleCanadien1rePartie_B.xml	1
Boulay_1923_CanadianSet1stChange_B.xml	1
Soucy_1928_LaJolieVeuve_A.xml	1
LaMadeleine_1927_QuadrilleFrancoAmericain4ePartie_C.xml	1
Boulay_1929_SetCanadianDeQuebec1rePartie_A.xml	1
Boulay_1929_SetCanadianDeQuebec2ePartie_B.xml	1
Allard_1929_Varsovienne_A.xml	1
LaMadeleineA_1929_QuadrilleDesSeigneurs2ePartie_A.xml	1
Name: Boulay_1923_GiguesPotPourri_1stTune_A.xml, Length: 627, dtype: object	

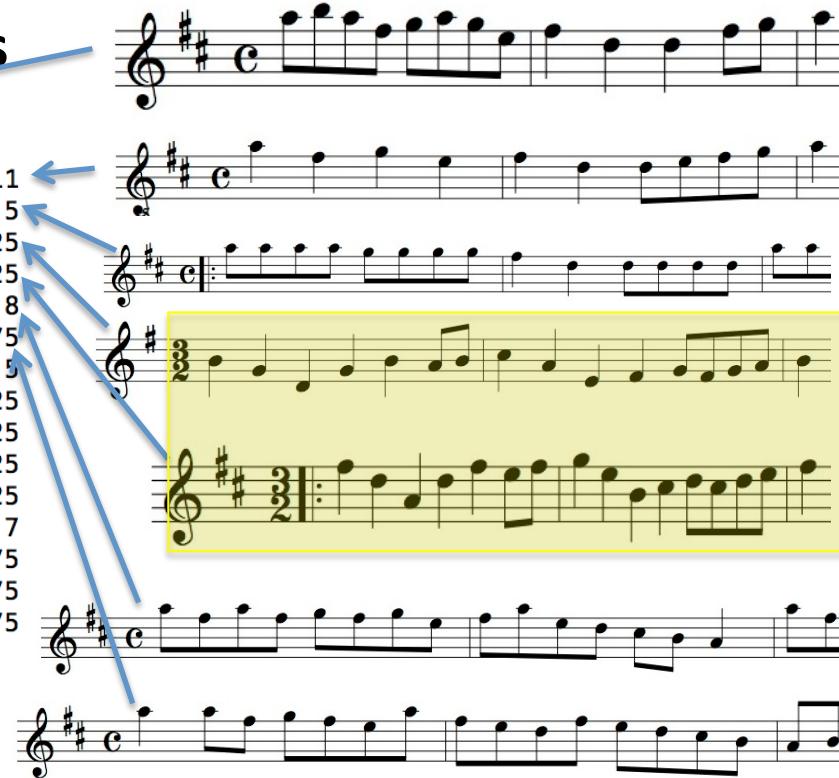
Boulay, Gigues pot-pourri, 1<sup>st</sup> tune (1923), A strain  
.xml')

Arthur-Joseph Boulay, Reel à quatre (1929), A strain

Boulay, Gigues pot-pourri, 1<sup>st</sup> tune (1923), B strain

# Rankings

>>> db.rankings_for("Soucy_1928_QuadrilleChampion2eFigure_A.xml")	
Allard_1931_ReelDeLaPolitique_B.xml	11
Blanchette_1937_ReelProvincial_B.xml	9.5
SoucyLafleur_1929_QuadrilleDesVieuxCanadiens1rePartie_B.xml	8.25
Soucy_1929_ReelEspagne_A.xml	8.25
Gauthier_1926_ReelOpera_C.xml	8
LaMadeleine_1928_QuadrilleDeMatane6ePartie_B.xml	7.75
LaMadeleine_1928_QuadrilleDesBucherons2ePartie_B.xml	7.5
Soucy_1927_LeCapitaineVoleur_A.xml	7.25
Gauthier_1926_ValseDeChezNous_B.xml	7.25
Allard_1929_GigueAmericain_B.xml	7.25
Soucy_1936_QuadrilleSteJulie_B.xml	7.25
Boulay_1929_ReelDesMatelots_B.xml	7
LaMadeleineA_1929_QuadrilleDesSeigneurs3ePartie_F.xml	6.75
Boulay_1923_CanadianSet2ndChange_A.xml	6.75
Soucy_1926_QuadrilleDeQuebec4ePartie_A.xml	6.75
...	
Soucy_1929_GigueDesTricoteuses_A-alt.xml	2
Allard_1928_LaMereBlanche_A.xml	2
Soucy_1929_ValseEmilia_A.xml	2
LarryOGaff_B.xml	2
LaMadeleine_1927_QuadrilleFrancoAmericain4ePartie_E.xml	2
LaMadeleine_1927_QuadrilleFrancoAmericain3ePartie_A.xml	2
Soucy_1926_AStMalo_B.xml	2
Allard_1929_BreakdownDuNuit_A.xml	2
Boulay_1923_CanadianSet1stChange_A.xml	2
Boulay_1924_CanadianSet1stChange_B.xml	2
Allard_1929_MarcheSirWilfridLaurier_C.xml	2
LaMadeleine_1927_QuadrilleFrancoAmericain3ePartie_B.xml	2
Boulay_1929_GigueDuBarbier_B.xml	1.75
Soucy_1926_SetAmericain3ePartie_A.xml	1.5
Soucy_1929_GigueDesAmoureux_B.xml	1.25
Name: Soucy_1928_QuadrilleChampion2eFigure_A.xml, Length: 627, dtype: object	



Problematic results in yellow.  
Possible solutions: divide ranking  
by length of the shorter  
fingerprint; only allow expansion  
when comparing fingerprints of  
different lengths; give certain  
strong beats more weight.

# Rankings

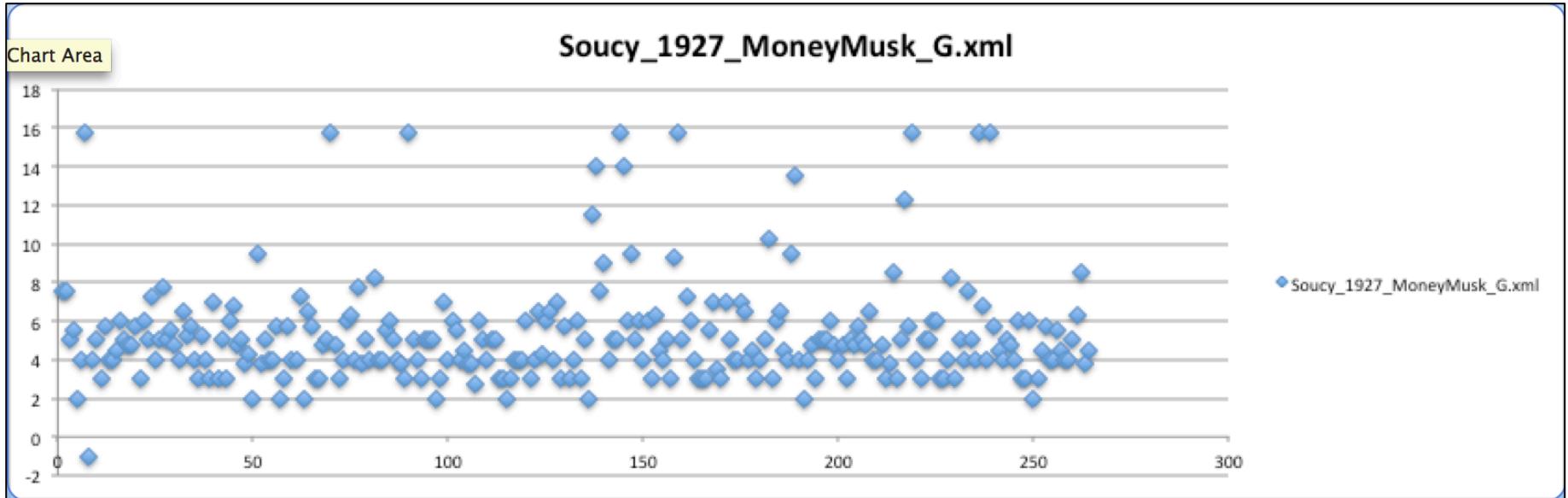
```
>>> db.rankings_for("Allard_1931_ReelDeLaPolitique_B.xml")
Soucy_1928_QuadrilleChampion2eFigure_A.xml
Allard_1929_ReelDesMontagnes_A.xml
Boulay_1925_SleepyHollow_B.xml
Soucy_1926_QuadrilleDeQuebec1rePartie_A.xml
Soucy_1929_GigueATiGus_D.xml
Soucy_1926_QuadrilleDeQuebec4ePartie_A.xml
Soucy_1927_QuadrilleDuCampagne4ePartie_A.xml
LaMadeleine_1928_QuadrilleDesBucherons2ePartie_B.xml
Gauthier_1926_PolkaCanadienne_B.xml
Boulay_1923_CanadianSet2ndChange_A.xml
LaMadeleine_1928_QuadrilleDeMatane6ePartie_B.xml
Blanchette_1937_ReelProvincial_A.xml
Blanchette_1937_ReelProvincial_B.xml
LaMadeleine_1928_Caillette_A.xml
Boulay_1929_SetCanadianDeQuebec2ePartie_B.xml
...
LaMadeleineA_1929_QuadrilleDesSeigneurs1rePartie_A.xml
Boulay_1929_ReelDesCinqMilles_B.xml
LaMadeleine_1927_QuadrilleFrancoAmericain3ePartie_B.xml
Allard_1933_ReelStadacona_A.xml
LaMadeleine_1927_QuadrilleFrancoAmericain5ePartie_B.xml
Allard_1929_ReelDuJournalier_A.xml
Boulay_1924_QuadrilleCanadien3ePartie_A.xml
Allard_1929_QuadrilleDeBeauharnois_A.xml
Gauthier_1926_ReelDuMarie_B.xml
LaMadeleine_1928_FrancoisMargotte_A.xml
Boulay_1929_ReelDuVercheres_B.xml
Soucy_1931_GigueDuPereLauzon_B.xml
Boulay_1924_QuadrilleCanadien4ePartie_A.xml
Soucy_1929_GigueDesTricoteuses_A-alt.xml
LaMadeleine_1927_QuadrilleFrancoAmericain1rePartie_A.xml
Name: Allard_1931_ReelDeLaPolitique_B.xml, Length: 627, dtype: object
```

Problematic results in yellow.  
Possible solutions: divide ranking  
by length of the shorter  
fingerprint; only allow expansion  
when comparing fingerprints of  
different lengths; give certain  
strong beats more weight.

## **Going forward**

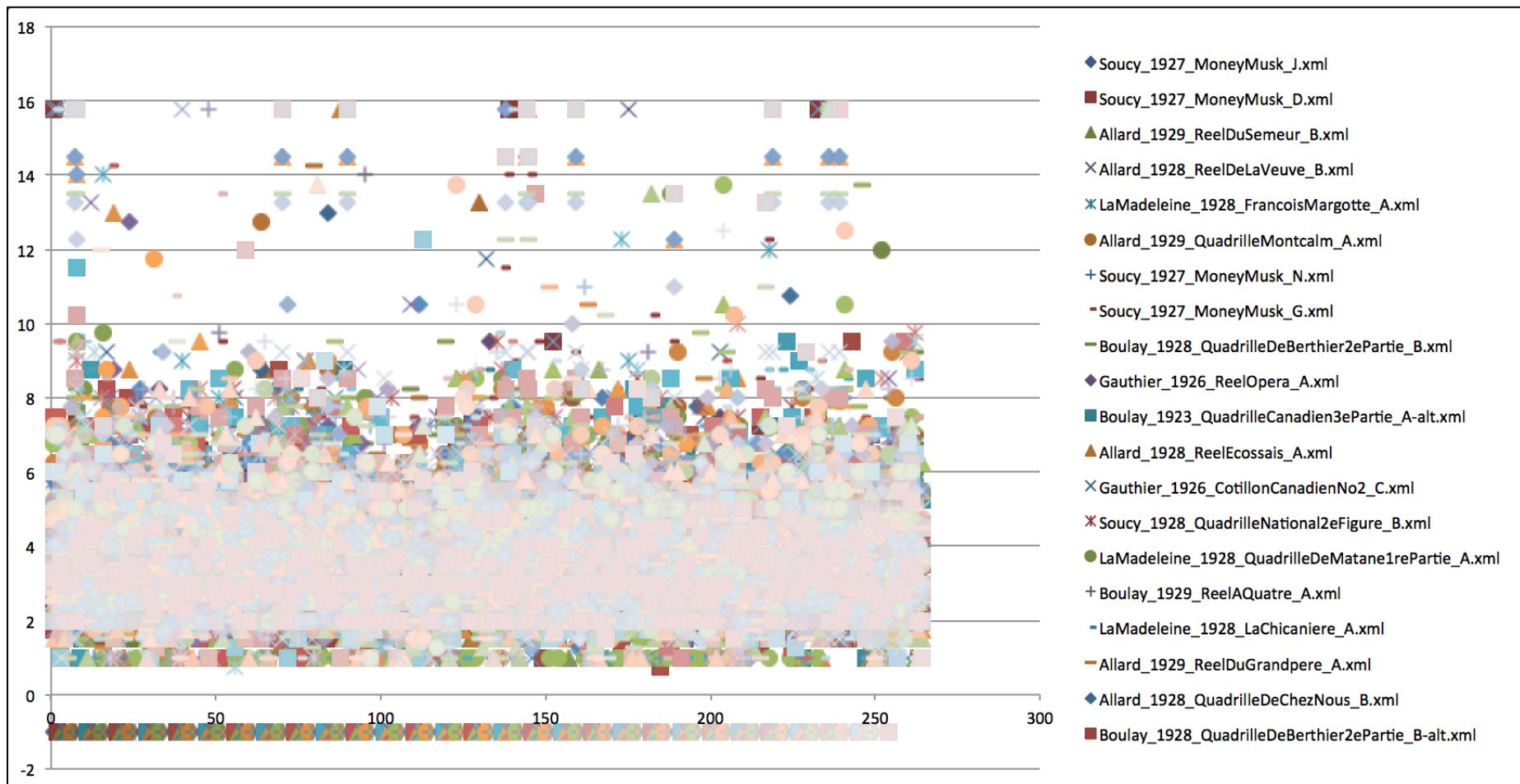
- Creating a visual representation
- Challenges
- Next steps

# Creating a visual representation



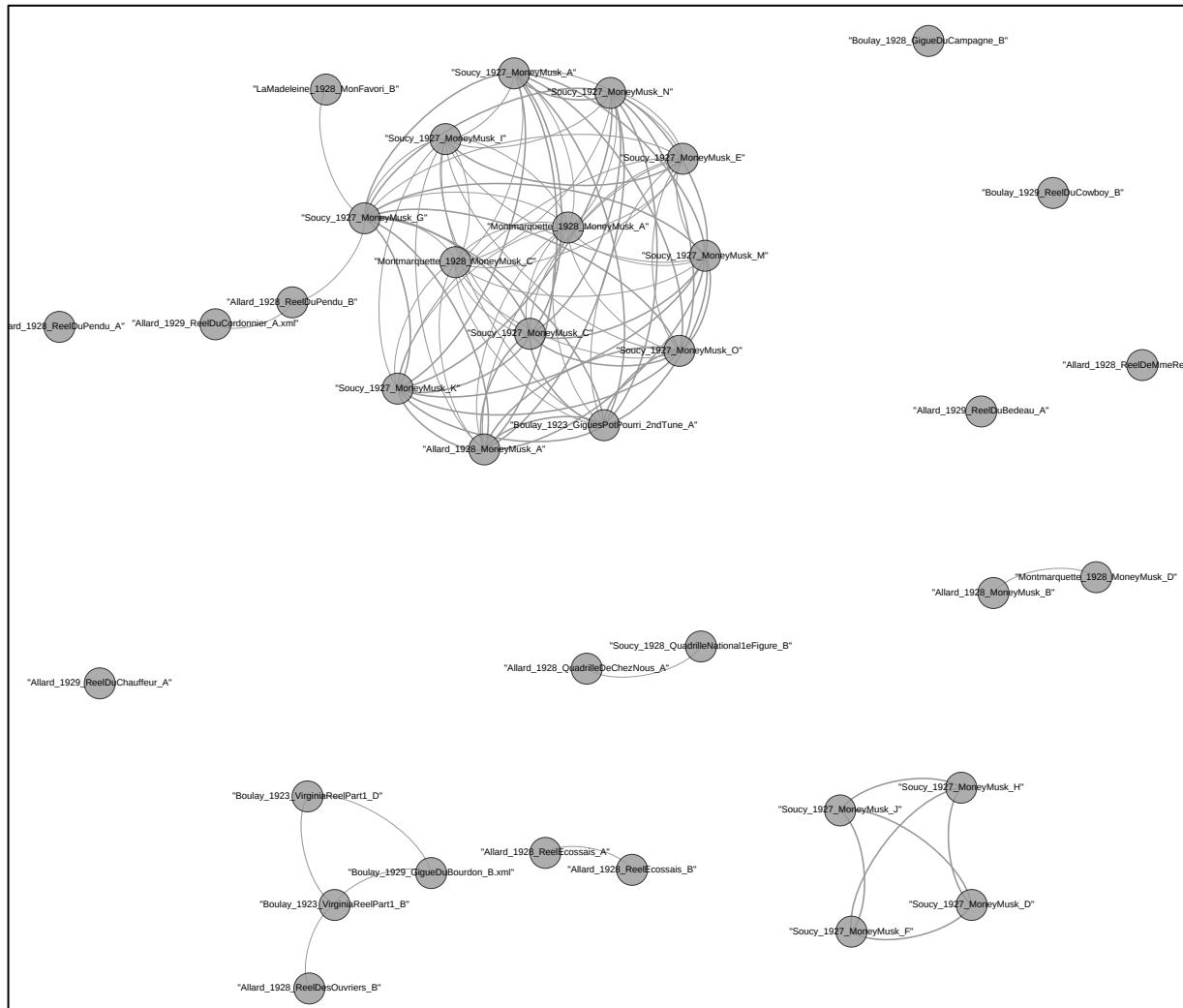
Sample visualization of similarity measures for a single strain. This partial database contains 264 strains. Graph created with Excel.

# Creating a visual representation



Sample visualization of a partial database of approx. 250 strains. Graph created using Excel. Each point represents a similarity comparison between two strains. Note that each comparison appears twice on this graph. A strain compared with itself results in a similarity measure of -1.

# Creating a visual representation



Sample visualization of 36 strains using Gephi (Yifan Hu layout). This mapping shows edges between all strains with a similarity measure of 10 or more. Edge thickness is proportional to similarity measure.

# Challenges

# Challenges

- Human error
  - Errors in transcription
  - Incorrect placement of the first two barlines in metrically irregular tunes
- Should the algorithm use more than the first two bars of a strain?
  - Perhaps combine this algorithm with a “first pass” algorithm based on edit distance?
- Weighting
  - S1 is weighted disproportionately because S9 is usually a repetition of S1
- Comparison
  - What if unequal truncations would give a better match? (see “human error” #2)
  - 4/4 vs. 6/8: perhaps add double-time settings of duple meter tunes?
  - Incorrect correspondences for fingerprints of different length: The correspondence function should only allow extension

# Comparing tunes in 4/4 and 6/8

The algorithm currently compares these two fingerprints as follows:

The image shows two musical staves. The top staff is in 4/4 time, G major, and the bottom staff is in 6/8 time, A major. Red arrows point from specific notes in the top staff to corresponding notes in the bottom staff, indicating a comparison or alignment of these notes.

Isidore Soucy, Quadrille Ste.-Julie (1936), A strain

Joseph Allard, La mère blanche (1928), A strain (transposed to A major)

# Comparing tunes in 4/4 and 6/8

However, this would be a better comparison.

The image shows two musical staves. The top staff is in 4/4 time, A major (two sharps), and the bottom staff is in 6/8 time, A major (two sharps). Both staves have a treble clef. The top staff is labeled "Isidore Soucy, Quadrille Ste.-Julie (1916), A strain". The bottom staff is labeled "Joseph Allard, La mère blanche (1928), A strain (transposed to A major)". Arrows point from specific notes in the top staff to corresponding notes in the bottom staff. A red double-headed arrow spans four measures of the top staff, with green arrows pointing from the first, third, and fourth measures to the corresponding measures in the bottom staff. Two purple arrows point from the fifth and sixth measures of the top staff to the corresponding measures in the bottom staff.

Possible solution: create 2/4 versions of all 4/4 fingerprints. This would effectively double the number of fingerprints in the database, however, and could potentially lead to many false positives.

# Comparing tunes in 4/4 and 6/8

Similarly, this is the current comparison for these tunes.

The image shows two musical staves side-by-side for comparison. The top staff is in 4/4 time, G major, and the bottom staff is in 6/8 time, D major. Both staves have a treble clef and two sharps in the key signature. The top staff is labeled "Isidore Soucy, Quadrille des bois, A strain". The bottom staff is labeled "Arthur Boulay, Quadrille de Berthier (1928), A strain (transposed to D major)". Red arrows point from the first note of the top staff to the first note of the bottom staff, indicating a correspondence between the first notes of each measure. A green arrow points from the last note of the top staff to the last note of the bottom staff, indicating a correspondence between the last notes of each measure. The music consists of eighth-note patterns.

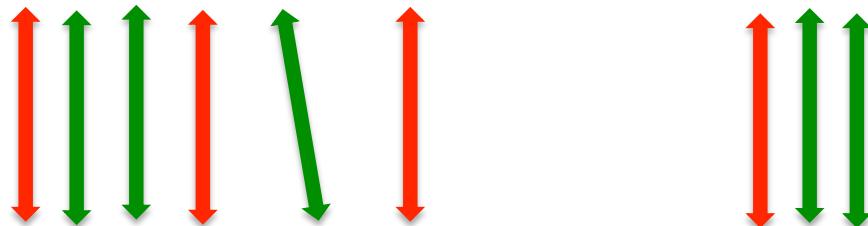
Arthur Boulay, Quadrille de Berthier (1928), A strain (transposed to D major)

# Comparing tunes in 4/4 and 6/8

However, this would be more accurate.



Isidore Soucy, Quadrille des bois, A strain



Arthur Boulay, Quadrille de Berthier (1928), A strain (transposed to D major)

# Incorrect correspondences for fingerprints of different length

The algorithm currently compares these two fingerprints as follows:

Comparison Results:		1	2	3	4	5	6
Strong Beat Comparison	[0.0, 0, 0]	[-1.5, 1, 1]	[0.0, 2, 2]	[0.0, 4, 6]	[-1.5, 5, 7]	[0.0, 8, 8]	
Displacement Comparison (Strong-Weak)	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Weak Beats Comparison (Matched Strong)	[1.0, 1.0]	[1.0, 1.0]	[1.0, 1.0]	[1.0, 1.0]	[1.0, 1.0]	[1.0, 1.0]	[1.0, 1.0]
Weak Beats Comparison (Mismatched Strong)	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Contour Comparison (Strong)	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Contour Comparison (Weak)	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Reversal Comparison (Strong)	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Reversal Comparison (Weak)	NaN	NaN	NaN	NaN	NaN	NaN	NaN

The image shows two musical staves. The top staff, labeled "Allard 1928, A strain", has a treble clef, a key signature of one sharp, and a common time signature. It contains six measures of music. The bottom staff, labeled "Duchesne 1934, D strain", has a treble clef, a common time signature, and a different rhythm pattern. Red arrows point from specific notes in the first measure of the top staff to notes in the second measure of the bottom staff, indicating a correspondence between them. The notes are eighth and sixteenth note patterns.

# Incorrect correspondences for fingerprints of different length

One of these would be a better comparison:

Comparison Results:		1	2	3	4	5	6
Strong Beat Comparison	[0.0, 0, 0]	[-1.5, 1, 1]	[0.0, 2, 2]	[0.0, 4, 6]	[-1.5, 5, 7]	[0.0, 8, 8]	
Displacement Comparison (Strong-Weak)	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Weak Beats Comparison (Matched Strong)	[1.0, 1.0]	[1.0, 1.0]	[1.0, 1.0]	[1.0, 1.0]	[1.0, 1.0]	[1.0, 1.0]	[1.0, 1.0]
Weak Beats Comparison (Mismatched Strong)	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Contour Comparison (Strong)	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Contour Comparison (Weak)	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Reversal Comparison (Strong)	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Reversal Comparison (Weak)	NaN	NaN	NaN	NaN	NaN	NaN	NaN

The image shows two musical staves. The top staff is in G major (two sharps) and the bottom staff is in G major (one sharp). Red arrows point from specific notes in the top staff to notes in the bottom staff, indicating potential correspondences. The text "Allard 1928, A strain" is above the top staff, and "Duchesne 1934, D strain" is below the bottom staff.

Solution: only allow “expansion” from a shorter fingerprint to a longer fingerprint

# Next steps

- Weighting
  - include variables for number of truncations and relative weight of certain strong beats
  - divide by length of shorter fingerprint
  - user-entered?
- Automate the preparation of the XML files (currently done by hand)
  - remove pickups
  - remove grace notes, chords, articulation markings
  - realize all repeats, including first and second endings
  - split file into strains and name accordingly
  - create 2/4 versions of 4/4 files?
  - double-stops? rests?
- Graphic display of results
- How might we adapt this algorithm to other repertoires?

Fingerprint 1:

	1	2	3	4	5	6	7	8	9
w	[-2.5]	[1.0, 1.5]	[2.5]	[5.0]	[-2.5]	[4.5]	[4.0]	[4.5]	[-2.5]
0.0	0	0	0	0	0	0	0	0	0
1.0	-1.5	-1	-1.5	4	-3.5	-0.5	-1	5	NaN
2.0	-2.5	-2.5	2.5	0.5	-4	-1.5	4	NaN	NaN
3.0	-4	1.5	-1	0	-5	3.5	NaN	NaN	NaN
4.0	0	-2	-1.5	-1	0	NaN	NaN	NaN	NaN
5.0	-3.5	2.5	2.5	4	NaN	NaN	NaN	NaN	NaN
6.0	-4	-3.5	2.5	NaN	NaN	NaN	NaN	NaN	NaN
7.0	-5	1.5	NaN	NaN	NaN	NaN	NaN	NaN	NaN
8.0	0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Thank you!

Questions? Comments? Suggestions?

Fingerprint 2:

	1	2	3	4	5	6	7	8	9
w	[-2.5]	[1.5]	[0.0]	[1.5]	[-2.5]	[1.5]	[-0.5]	[-1.0]	[-2.5]
0.0	0	0	0	0	0	0	0	0	0
1.0	-1.5	-1	-1.5	4	-4	0.5	-1.5	5	NaN
2.0	-2.5	-2.5	2.5	0	-4	0.5	-1.5	5	NaN
3.0	McGill								
4.0	0	-2.5	-1	-1	0	NaN	NaN	NaN	NaN
5.0	-4	-2	-2.5	4	NaN	NaN	NaN	NaN	NaN
6.0	-3.5	-3.5	2.5	NaN	NaN	NaN	NaN	NaN	NaN
7.0	-5	1.5	NaN						
8.0	0	NaN							



McGill



Schulich School of Music  
École de musique Schulich