MELODIC SIMILARITY IN TRADITIONAL FRENCH-CANADIAN INSTRUMENTAL DANCE TUNES

Laura Risk **Lillio Mok** **Andrew Hankinson** **Julie Cumming**

Schulich School of Music
Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT)
McGill University
laura.risk@mail.mcgill.ca

ABSTRACT

Commercial recordings of French-Canadian instrumental dance tunes represent a varied and complex corpus of study. This was a primarily aural tradition, transmitted from performer to performer with few notated sources until the late 20th century. Practitioners routinely combined tune segments to create new tunes and personalized settings of existing tunes. This has resulted in a corpus that exhibits an extreme amount of variation, even among tunes with the same name. In addition, the same tune or tune segment may appear under several different names.

Previous attempts at building systems for automated retrieval and ranking of instrumental dance tunes perform well for near-exact matching of tunes, but do not work as well in retrieving and ranking, in order of most to least similar, variants of a tune; especially those with variations as extreme as this particular corpus. In this paper we will describe a new approach capable of ranked retrieval of variant tunes, and demonstrate its effectiveness on a transcribed corpus of incipits.

1. INTRODUCTION

Commercial recordings of French-Canadian instrumental dance tunes from the 1920s through the 1980s document a working-class repertoire now celebrated as the traditional instrumental music of Québec [19]. However, the musical contents of these recordings remain largely unexamined. The only detailed musicological study is limited to a subset of metrically irregular tunes [8]. In this paper we outline the challenges associated with studying this repertoire and describe a new system developed to aid in finding and ranking similarity between tunes in this repertoire. This system was built in response to two musicological challenges: to determine the degree of shared repertoire among early commercial recording artists in Montréal, and to identify how a single tune is varied in different renditions. Using our system we have identified a number of concordant tunes (versions of the same tune) previously unrecognized as being musically related.

Broadly speaking, the traditional instrumental music of Québec is similar to the instrumental traditions of Ireland, Scotland and the United States. The repertoire consists primarily of short, fast-paced dance tunes usually performed on the violin, accordion, or harmonica. With very few exceptions, each tune has at least two strains (sections), commonly labeled "A" and "B." Many of the tunes have their roots in British Isles and American fiddling traditions, though others are derived from popular French songs or early twentieth-century marching-band repertoire [8; 13; 23].

2. THE CHALLENGE

The French-Canadian tradition has developed almost exclusively as an aural and recorded tradition. With few notated sources, musicians would often learn tunes "on the fly," constructing their own versions from memory and injecting their own personal style. From the 1930s through the 1960s radio broadcasts played a significant role in aural transmission of this repertoire. One musician recalled, "I would listen to the radio with my brother, and afterwards we would sing the melodies in our room. We spent our time constantly asking ourselves if it was really correct" [9]. As a result of this mode of transmission, and in the absence of a culture of "correctness" [7], many tunes performed and recorded in Québec exist in multiple, equally valid settings. Tunes would be modified by transposing all or part of the tune, reworking melodic figurations, adding and subtracting beats, composing new strains, or combining strains from several tunes to form new tunes. This diversity of interpretation is clearly documented on commercial recordings from the era.

Tune titles were often lost or altered in transmission. Fiddler Yvon Mimeault, for instance, assigned his own titles to most of the tunes that he learned from the radio in the 1950s [14]. Other musicians renamed tunes quite intentionally. In the 1920s and 30s, fiddler Isidore Soucy sometimes recorded a tune for one record label under one title, and within months recorded the same tune for a different record label under a different title. The tune "Money Musk" is an exception. Between 1920 and 1980 it was recorded over twenty times, frequently with significant melodic and rhythmic variation and

additional strains, but almost always under the same title.

To illustrate with one example, fiddlers Isidore Soucy, Joseph Ovila LaMadeleine, and Joseph Allard all released settings of the same melody in 1928. Their recordings were titled, respectively, "Reel du bon vieux temps," "Reel princesse," and "Reel de Mme. Renault" (note that Soucy reverses the order of the A and B strains with respect to LaMadeleine and Allard). Although these are recognizably the same melody, they have a significant degree of melodic and metrical variation (figure 1). Audio files for all three are available through the Virtual Gramophone website of Library and Archives Canada [20].

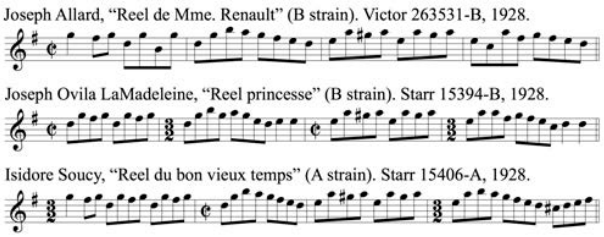

Figure 1: Incipits for three variants of a strain recorded in 1928.

The earliest recording sessions of traditional instrumental music in Québec were quick and largely unrehearsed. A pianist or guitarist would usually accompany a soloist with minimal rehearsal time prior to recording. Some of the recordings also include accompaniment on jaw harp (*guimbarde*) or spoons. The performances are unedited and were usually completed on either the first or second take. Some contain obvious musical errors, such as missed entries or wrong notes.

These performances were pressed to 78 RPM records, with one three-minute rendition of a tune per side [25]. These recordings contain a significant amount of background noise introduced in the recording chain, along with the standard problems of the 78 RPM format such as hiss, pops, and clicks.

The noisy recording environment and the relatively poor quality of the recordings result in recordings that are difficult to follow, even for human listeners. Due to these difficulties we decided not to explore signal–based approaches to analyzing this repertoire. Instead, our approach was to: 1) transcribe the A and B strains of a recording into MusicXML using a notation editor (Finale), and 2) devise a system for analyzing and computing the distance between two variants of the same tune.

Duval [8] estimates that the traditional instrumental music of Québec contains at least 5000 distinct tunes, not including variants. This gives a potential corpus of well over 10,000 strains. We are currently using our system to parse a database containing 710 strains. Of these, 667 were recorded between 1923 and 1929 and 59 strains are from renditions of "Money Musk" (16 strains of "Money Musk" were recorded between 1923 and 1929). This collection contains approximately 85% of all French-Canadian recordings of traditional instrumental music on the violin prior to 1930, and approximately 50% of those recorded on any instrument prior to 1930. This selection of repertoire is clearly not random, but rather reflects the imperatives of several musicological questions, as discussed below.

3. PREVIOUS WORK

Scholars of aural traditions have long been fascinated by repertoire variation, and comparative studies abound. Bayard [2] proposed an influential theory of "tune families" by which the bulk of British Isles and North American folk song melodies could be categorized as variants of a small number of distinct prototypical melodies. Cowdery [4], drawing examples from Irish traditional music, pointed out that musicians do not think in terms of abstract prototypes but rather create new tunes or variants by reworking and combining segments of known repertoire. He argued that tune families were more appropriately defined by the presence of recurring melodic motives, and not by their degree of deviation from a "standard" or "ideal" version of the tune.

Our query and ranking system is intended to help scholars study the diversity of melodic variants within a given corpus. Musically, our approach is similar to the approach described in three previous studies. Ó Súilleabháin [22] analyzed the melodic variations of Irish fiddler Tommie Potts according to a framework of "set accented tones." Goertzen [10] argued that seemingly disparate variants of Texas contest-style fiddle tunes are linked to a shared sense of each tune's "essence," itself composed of tune-specific musical markers. Duval [8] analyzed temporal variation as an innovative element of performance practice in French-Canadian tunes. However, none of these studies were performed using computational tools.

Several existing online resources allow users to search fiddle tunes by musical incipit. Both the Scottish Music Index [12] and the Traditional Tune Archive [18] classify tunes according to numerical theme codes that contain the scale degrees of the strong beats of the first two bars. Users may search on these sites for theme codes that exactly match a given string and that begin with that string, but may not search for tune variants.

TunePal [6] is a tool that translates audio into symbolic notation (ABC) and then compares that notation to a crowd-sourced database of traditional Irish tunes using an edit-distance function. TunePal looks for exact or similar strings of ABC regardless of metrical placement and is effective at identifying tunes and tune settings with a small amount of melodic variation (provided there has been no transposition). However, variants with significant melodic variation may not be considered a match.

Van Kranenburg [26] presents a comprehensive survey of computational modelling of similarity to Dutch folk songs. He concludes that identifying characteristic motifs is the most important factor when determining similarity between two melodies. As well,

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
w	[nan]	[0.5]	[2.5]	[4.0]	[2.5]	[-1.0]	[-0.5]	[-1.0]	[2.5]	[0.5]	[2.5]	[0.5]	[-2.0]	[-1.5]	[-0.5]	[-1.0]	[nan]
0.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1.0	-0.5	-2	-1.5	1.5	4.5	-2	-1.5	0	2	-2	2	-2	2.5	-1	-1.5	1.5	NaN
2.0	-2.5	-3.5	0	0	2.5	-3.5	-1.5	2	0	0	0	0.5	1.5	-2.5	0	NaN	NaN
3.0	-4	-2	4.5	4	1	-3.5	0.5	0	2	-2	2.5	-0.5	0	-1	NaN	NaN	NaN
4.0	-2.5	2.5	2.5	2.5	1	-1.5	-1.5	2	0	0.5	1.5	-2	1.5	NaN	NaN	NaN	NaN
5.0	2	0.5	1	2.5	3	-3.5	0.5	0	2.5	-0.5	0	-0.5	NaN	NaN	NaN	NaN	NaN
6.0	0	-1	1	4.5	1	-1.5	-1.5	2.5	1.5	-2	1.5	NaN	NaN	NaN	NaN	NaN	NaN
7.0	-1.5	-1	3	2.5	3	-3.5	1	1.5	0	-0.5	NaN	NaN	NaN	NaN	NaN	NaN	NaN
8.0	-1.5	1	1	4.5	1	-1	0	0	1.5	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
9.0	0.5	-1	3	2.5	3.5	-2	-1.5	1.5	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
10.0	-1.5	1	1	5	2.5	-3.5	0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
11.0	0.5	-1	3.5	4	1	-2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
12.0	-1.5	1.5	2.5	2.5	2.5	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
13.0	1	0.5	1	4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
14.0	0	-1	2.5	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
15.0	-1.5	0.5	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
16.0	0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Figure 2: Interval matrix for the B strain of "Reel de Mme. Renault" (Joseph Allard, 1928; see figure 1). The column headings (1 to 17) indicate the strong beats. The row headings (0.0 to 16.0) indicate the metric offset from a given strong beat. The first row (w) gives the weak beats for each strong beat as intervals from that strong beat. Interval values indicate number of whole steps.

he demonstrates that the frequency of global features (interval features and other pitch-based features, duration-ratio features, and other rhythmic features) is not sufficient to identify similar melodies, but that searching for ordered sequences of certain features at a local level may help locate similar song melodies. The system described in this paper builds on these findings.

4. METHODOLOGY

The design of our system exploits some general characteristics of French-Canadian instrumental dance tunes. We determine an optimal alignment between two four-measure incipits and then compare certain features at corresponding locations in the incipits. Transposition is common in this repertoire and tonal center is sometimes ambiguous, so all tunes are internally represented as a melodic contour. Each tune has two or more strains that function more or less independently and must therefore be treated separately. Most strains may be uniquely identified by a four-bar incipit. Since most tunes repeat after four measures, a transcribed incipit contains the primary motivic material for that strain. As in related British Isles and North American fiddle repertoires, metrical placement matters: the notes that fall on the strong beats are more essential to the identity of a strain than those on the weak beats [11].

Our system requires a set of strains in symbolic format and operates in two phases. A construction phase is used to build matrix representations of each strain's incipit, after which a comparison phase computes the pairwise similarity between matrices. The construction phase first scans and truncates the prepared strains to a four-measure incipit by using the `music21` toolkit [5]. Melodic, or horizontal, intervals between the first note on the first strong beat and the notes on all other strong beats are then identified using the VIS analysis framework [1]. Horizontal intervals between each note on a weak beat and the note on its preceding strong beat are also indexed and stored. The resulting feature vectors of these pair-wise indexed intervals create an interval matrix. Each *j*th column of the matrix represents the horizontal intervals of the melody between the *j*th strong beat and all subsequent strong beats. Figure 2 shows the interval matrix for the first strain shown in figure 1.

The comparison phase aligns the strong beats of two interval matrices before computing feature similarity. For matrices of the same length, we follow earlier musicological studies of related British Isles traditions [11; 22] and align the strong beats of the incipits bijectively (beat-to-beat). In other words, the idiomatic alignment for two equal-length incipits *A* and *B* is the *i*th beat in *A* to the *i*th beat in *B*. For matrices of different lengths, a standard longest-common-subsequence (LCS) dynamic programming algorithm [17; 21] is used to find the best alignment between two incipits considering the note on each strong beat.

After alignment this phase iteratively matches the notes of each aligned strong and weak beat in the two matrices. Non-matching notes are checked for three possible musical variation techniques: displacement, reversal and contour similarity. The aligned pair of strong beats (*i, j*) is non-matching when the note on beat *i* in *A* is not the same as the note on beat *j* in *B*. Thus, given two incipits *A* and *B* and each *k*th pair (*i*, *j*) of aligned but non-matching strong beats, *displacement* of notes on strong beats to a corresponding weak beat occurs when the note on the *i*th beat is amongst the notes in the weak beats after the *j*th beat, or vice versa. A *reversal* of notes on strong beats occurs when the note on the *i*th beat is the same as that on beat $j + 1$, and vice versa. Finally, *contour similarity* is detected when the horizontal interval between the notes on beats *i* and $i + 1$ are the same as the interval between beats *j* and $j + 1$.

To account for cases where the incipits may otherwise match but have different notes on the first aligned strong beat, the algorithm moves to the next column of both interval matrices and repeats the comparison phase. This recursive strategy is only used on up to half the interval columns of the incipits. For each repetition of the comparison phase, a similarity matrix is constructed from the results of analyzing the matching, displacement, reversal, and contour similarity of the aligned strong beats in the corresponding interval columns between incipit pairs.

Each row of a similarity matrix corresponds to the

Best Result:

	1	2	3	4	5	6
Strong Beat Comparison	[0.0, 0, 0]	[-2.0, 1, 1]	[nan, 2, 2]	[-2.0, 3, 3]	[2.5, 4, 4]	[nan, 5, 5]
Displacement Comparison (Strong-Weak)	NaN	NaN	0	NaN	NaN	1
Weak Beats Comparison (Matched Strongs)	[1.0, 1.0]	[1.0, 1.0]	NaN	[1.0, 1.0]	[0, 0]	NaN
Weak Beats Comparison (Mismatched Strongs)	NaN	NaN	[1.0, 1.0]	NaN	NaN	[0, 0]
Contour Comparison (Strongs)	NaN	NaN	NaN	NaN	NaN	NaN
Contour Comparison (Weaks)	NaN	NaN	NaN	NaN	NaN	NaN
Reversal Comparison (Strongs)	NaN	NaN	NaN	NaN	NaN	NaN
Reversal Comparison (Weaks)	NaN	NaN	NaN	NaN	NaN	NaN
Shorter Incipit Length	17	17	17	17	17	17
Longer Incipit Length	17	17	17	17	17	17
Number of Truncations	1	1	1	1	1	1

Best Similarity Measure: 63.1067961165

Figure 3: The similarity matrix comparing the B strain of "Reel de Mme. Renault" (Joseph Allard, 1928) and the B strain of "Reel princesse" (Joseph Ovila LaMadeleine, 1928). See figure 1 for incipits. This figure shows the feature values for the first six matched strong beats in the two incipits.

result of a particular feature analysis, and the columns represent the *k*th pair of aligned strong beats from each incipit's interval matrix. Each cell in the similarity matrix is thus the result of a feature analysis for a given strong beat pair between incipits *A* and *B*. Entries in each similarity matrix are then combined with a weighting factor to yield a single similarity measure for each similarity matrix (figure 3). The "Strong Beat Comparison" matrices indicate the value of matching strong beats and identify aligned strong beats. The "Weak Beats Comparison" matrices indicate the fraction, and relative order, of matching weak beats.

Our weighting scheme is determined by trial and error. Up to 85% of the weight value is assigned to matching strong beats, displaced or reversed strong beats, and matching contour, with the remainder used for weighting matching weak beats. This weighting scheme has the effect of selecting incipits with a high percentage of matching strong beats and then ranking those selected according to the number of matching weak beats.

5. RESULTS

Two outputs from this system may be useful to musicians, musicologists, and other researchers: the similarity matrices, which allow users to directly compare two incipits, and a ranked list of similarity measures between one strain and all other strains in the database.

We attempted a traditional precision and recall analysis but found it to be an unsuitable measure of effectiveness. It returned unnaturally high results because the weighting was determined to maximize precision and recall for known concordant strains. As noted in the discussion below, precision and recall for 25 variants of the A strain of "Money Musk" were either 100% (n=10) or 96% (n=15), given 24 relevant items, 24 retrieved items, and a database of 59 items.

Figure 4 gives incipits for the top 5 strains in the ranked list for the B strain of Joseph Allard's "Reel de Mme. Renault," as compared to 666 other strains recorded between 1923 and 1929. A human-provided musical analysis identified strains 1 and 3 ("Reel du bon vieux temps," "Reel princesse") as the only concordances in the database.

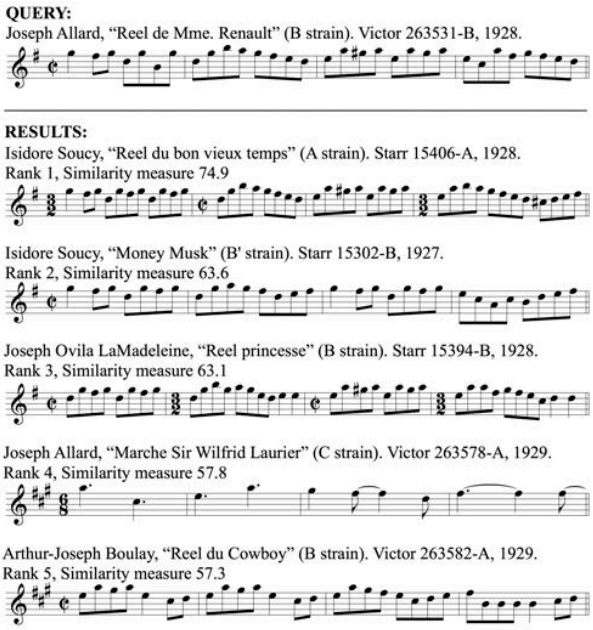

Figure 4: Similarity results for a query of the B strain of "Reel de Mme. Renault" (Joseph Allard, 1928)

To test our system, we randomly selected a test set of 100 strains from the full database of 710 strains. We then selected query strains at random from within this test set. We confirmed via a human-supplied musical analysis that each query strain had at least one concordance in the test set. Those that did not were discarded and new strains were randomly selected, until we reached a total of 10. Approximately 50% of the strains in the full database are not concordant with any other strains in the database.

We compared each of these query strains to the test set using four different ranking approaches (figure 5). The Levenshtein and Geometric Distance measures were drawn from the similarity evaluation system described in [16]. The MATT2 system [6] is designed for a repertoire of Irish tunes and is the search algorithm underlying the TunePal app described earlier. It presumes that transcriptions of tunes on "unusually pitched" instruments (instruments with repertoire-specific and non-standard tunings) have been normalized to a single fundamental pitch ("transposition invariance").

QUERY: Performer, year, tune title (strain); number of concordant strains (CS) in test set.	• *How many concordant strains (CS) are in the top 10 results for the query strain?* • *What are the rankings of those CS?* *Note: CS were determined by human-provided analysis*			
	Our system	MATT2 (fundamental pitch normalized)	Levenshtein Distance	Geometric Distance
Allard, 1928, "Reel du Pendu" (B); 2 CS	2 CS; rank 1, 2; SM 77, 69.	2 CS; rank 1, 2.	0 CS.	0 CS.
Allard, 1929, "Reel des Violoneux" (B); 2 CS	2 CS; rank 1, 2; SM 89, 75.	2 CS; rank 1, 3.	2 CS; rank 1, 2	2 CS; rank 1, 5.
Boulay, 1923, "Gigues Pot-Pourri" (2nd tune, A); 4 CS	4 CS; rank 1, 2, 3, 4; SM 91, 91, 90, 80.	4 CS; rank 1, 2, 3, 4.	0 CS.	1 CS; rank 6.
Duchesne, 1938, "Money Musk Americain" (A); 4 CS	4 CS; rank 1, 2, 3, 4; SM 93, 91, 90, 75.	4 CS; rank 1, 2, 3, 4.	0 CS.	1 CS; rank 2.
Joyal, 1956, "Money Musk" (B); 3 CS	3 CS; rank 1, 2, 7; SM 80, 76, 46.	2 CS; rank 2, 3.	1 CS; rank 1.	1 CS; rank 1.
Lajoie, 1951, "Money Musk" (B); 3 CS	3 CS; rank 1, 2, 3; SM 80, 67, 55.	2 CS; rank 1, 2.	0 CS.	0 CS.
LaMadeleine, 1927, "Quadrille Franco-Americain 6e partie" (A); 1 CS	1 CS; rank 1; SM 100.	1 CS; rank 1.	1 CS; rank 1.	1 CS; rank 1.
Potvin, 1980, "Money Musk" (A'); 4 CS	4 CS; rank 1, 2, 3, 4; SM 93, 91, 86, 75.	4 CS; rank 1, 2, 3, 4.	2 CS; rank 4, 6.	0 CS.
Soucy, 1925, "Gigues irlandaises no. 2" (2nd tune, A); 2 CS	2 CS; rank 1, 2; SM 91, 49.	2 CS; rank 1, 2.	1 CS; rank 1.	2 CS; rank 1, 2.
Soucy, 1927, "Quadrille Laurier, 3e partie" (A); 1 CS	1 CS; rank 1; SM 60.	0 CS.	1 CS; rank 1.	1 CS; rank 1.

Figure 5: Query results for 10 strains out of a test set of 100 strains.

As expected, all of these analytical systems performed well when identifying exact or near-exact matches. However, our system was able to identify more extreme melodic and metrical variants and to supply a ranking of those variants. Our system also identified concordances in transposed keys.

6. DISCUSSION

We built this system in response to two musicological challenges. First, we wanted to identify concordances in the earliest commercial recordings of French-Canadian tunes in order to determine the degree of shared repertoire among these recording artists. Second, we wanted to analyze variation technique in a single tune, "Money Musk," due to its popularity in recordings of the era.

By applying our approach to a database of 667 strains recorded between 1923 and 1929, we were able to identify nearly 150 concordant strains. Most of these concordances were previously unrecognized as related tunes.

Using these results in combination with archival research, we have been able to identify patterns of musical borrowing for certain musicians. Of the 16 sides that fiddler Isidore Soucy recorded for Columbia Records in New York City in 1927–1929 [24], for instance, eight were tunes that he had recorded for the Starr label in Montréal only a few months earlier, most under different titles. In contrast, when Soucy borrowed from his Starr releases for other Starr recordings, he usually re-recorded only single strains (combined with new material or with strains borrowed from other tunes).

We have been able to document the musical links between a small group of fiddlers living and working in the Montréal region in the late 1920s. These musicians often re-recorded the same tunes and strains within weeks of each other. Joseph Allard and Isidore Soucy, for example, recorded the same tune in late 1928 under the titles "Quadrille Acadien" and "Gigue Indienne," respectively (Victor 263543–A, Starr 15517–A). In the summer of 1927, Willie Ringuette and Isidore Soucy added the same C strain to two different tunes (Starr 15347–A, Starr 15363–B).

Finally, we were able to identify French-Canadian variants of many common North American tunes such as "Soldier's Joy," "Haste to the Wedding," "Fisher's Hornpipe," "Bristol Hornpipe," "Rickett's Hornpipe," "Chicken Reel," "Irish Washerwoman," "Keel Row," "Lord McDonald" and "Home Sweet Home."

We applied our system to a database of 59 strains drawn from 13 renditions of "Money Musk." All "Money Musk" settings include some version of two particular strains, usually labeled A and B, though not all performers play the A strain first and the B strain second. These strains may be easily recognized: both begin with a down-and-back motion that outlines a tonic chord, though the A strain begins on the fifth scale degree and the B on the first scale degree. (For three early and quite varied renditions of "Money Musk," listen to recordings on the Virtual Gramophone by Isidore Soucy [Starr 15302–B, 1927], Joseph Allard [Victor 263527–B, 1928] and Alfred Montmarquette [Starr 15475–A, 1928]).

A human-supplied musical analysis of these 59 strains identified 25 variants of the A strain, 15 variants of the B strain, 10 strains that were neither A nor B variants, and 9 strains that could be conceived of as distant variants of A (4 strains) or B (5 strains). In addition, this analysis revealed two types of B strains: those with an ascending melodic contour in the second bar (7 strains), and those with a descending figure (8 strains). We used our system to generate ranked lists for each of the 59 strains. For the 25 A strains, the top 24 results in the ranked list contained either 24 (n=10) or 23 (n=15) of the remaining A strains. The B-strain results were more complex and are summarized in figure 6.

These results suggest that the A-strain variants of "Money Musk" are more similar to each other than are the B-strain variants, and that B-ascending variants are more diverse than B-descending. This analysis also allows us to identify certain variants as musical outliers. As noted above, 15 of the A strains recalled 23 of 24 other A strains. In 12 of these instances, the missing strain was the same. This suggests that this strain would be a good candidate for further study.

The results in figure 6 also point to a split in the B-ascending strains, between those that are most similar to

the other B-ascending strains and those that are equally similar to B-ascending and B-descending strains. In particular, the strain recorded by Arthur-Joseph Boulay stands out for its dissimilarity to other B-ascending strains. These B-strain results suggests that the original musicological analysis that classified the B strains as either ascending or descending may need to be refined with reference to additional significant features.

B-ascending strains (n=7) for the tune "Money Musk"

Artist, year (strain)	Number of other B strains (n=14) in top 14 items on ranked list. This includes both B-ascending and B-descending strains.	Number of other B-ascending strains (n=6)...	
		...in top 14 items on list	...in top 6 items on list
W. Boivin, 1974 (B)	13	5	1
W. Boivin, 1974 (B')	13	5	2
W. Boivin, 1974 (B'')	6	6	4
A.-J. Boulay, 1923 (B)	8	0	0
U. Potvin, 1980 (B)	7	5	5
U. Potvin, 1980 (B')	12	5	5
I. Soucy, 1927 (B)	9	5	5

B-descending strains (n=8) for the tune "Money Musk"

Artist, year (strain)	Number of other B strains (n=14) in top 14 items on ranked list (includes B-ascending and B-descending).	No. of other B-descending strains (n=7)...	
		...in top 14 items on list	...in top 7 items on list
J. Allard, 1928 (B)	11	7	6
G. Joyal, 1956 (B)	11	7	7
G. Joyal, 1956 (B')	11	7	7
G. Lajoie, 1951 (B)	12	7	5
G. Lajoie, 1951 (B')	10	7	5
A. Montmarquette, 1928 (B)	10	7	6
É. Picard, 1930 (B)	10	7	5
A. Richard, 1975 (B)	11	7	5

Figure 6: Results for 15 B strains of "Money Musk" out of a database of 59 strains.

Examples such as these suggest that our approach may help scholars of instrumental dance music achieve a more nuanced study of musical similarity. Specifically, our system may help to identify concordances, parse degrees of melodic variation, and pinpoint instances that require further examination. The system also provides tools—the similarity matrices and the ranked lists—to facilitate such examination.

Certain instances of comparison remain problematic, however. The system does not currently recognize changes in meter, occasionally resulting in incipits that are slightly shorter or longer than four measures. This is the case for both "Reel princesse" and "Reel du bon vieux temps" (figure 1). The ranked list results for such cases are still reasonably accurate (figure 4). Note also that the placement of the barlines in metrically irregular renditions of tunes is at the discretion of the transcriber.

In addition, the French-Canadian repertoire contains some tunes with variants in both compound meter (9/8 or 6/8) and simple meter (3/2, 4/4, 2/2, or 2/4). In such cases, the system does not always find a satisfactory alignment between strong beats.

The system may also generate incorrect results when the two incipits are of different lengths. This is largely because variations in length of musically similar strains are due to an expansion of the shorter to the longer. While a naïve dynamic programming approach to alignment is insensitive to expansion, this issue may be solved by reducing the weighting on alignments that compress the shorter incipit.

7. FUTURE WORK

Although our system is currently designed for the specific attributes of French-Canadian fiddle tunes, the comparison functions and weighting calculation may be adapted for other repertoires. Our system may be particularly useful for repertoires in which new melodies are constructed using modified segments of extant melodies. Such repertoires are primarily aural, but may also include notated repertoires such as Renaissance Masses based on pre-existent material. This would require modifying the system for polyphonic sources. More immediately, we would like to investigate the application of our system to British Isles and North America fiddling traditions. We do not anticipate needing to revise the comparison functions and weighting calculation for this repertoire, and thousands of these tunes are already available in symbolic notation via online databases [3; 15; 18].

Our system identifies nuances between two strains and is particularly useful for identifying strains with a high degree of variation. However, we recognize that our approach may be less versatile than a more generalized comparison function such as an edit distance or Earth Mover's Distance function. Eventually we may seek to combine our system with a "first pass" edit distance or Earth Mover's Distance function.

8. CONCLUSION

Musical repertoires that circulate primarily in aural tradition often contain significant variance between different instances of the same tune. Analyzing variation and transformation in such repertoires has been an important part of ethnomusicology, musicology and folklore scholarship for decades. This paper has presented a novel tool to aid researchers in variance analysis in instrumental dance tunes.

The source code for our system has been published under an open source license, available on GitHub at http://github.com/ELVIS-Project/fiddle-tunes.

We believe that our system may be of practical use for musicologists and musicians specializing in the traditional instrumental musics of the British Isles and North America. It may also prove a useful model when building analytical tools for other repertoires containing a large number of variants.

ACKNOWLEDGEMENTS

The authors would like to thank the following people for their help and assistance: Ichiro Fujinaga, David Brackett, Bryan Duggan, Matt Kelly, Peter van Kranenburg, Dorothea Blostein, Marc Bolduc, and Jean Duval. This work was supported by the Social Sciences and Humanities Research Council of Canada as part of the Single Interface for Music Score Searching and Analysis (SIMSSA) Project, and by a Doctoral Award from Bibliothèque et Archives nationales du Québec.

9. REFERENCES

[1] Antilla, C., and J. Cumming: "The VIS Framework: Analyzing counterpoint in large datasets." In *Proc. Conf. Int'l Society for Music Information Retrieval*. Taipei, 71–6. 2014.

[2] S. Bayard: "Prolegomena to a study of the principal melodic families of British-American folk song." *The Journal of American Folklore* 63 (247) 1–44, 1950.

[3] J. Chambers: "JC's ABC tune finder." 2015. http://trillian.mit.edu/~jc/cgi/abc/tunefind (accessed 3 May 2015).

[4] J. Cowdery: "A fresh look at the concept of tune family." *Ethnomusicology* 28 (3) 495–504, 1984.

[5] Cuthbert, M., and C. Ariza: "music21: A toolkit for computer-aided musicology and symbolic music data." In *Proc. Conf. of the Int'l Society for Music Information Retrieval*. Utrecht, The Netherlands, 637–42. 2010.

[6] B. Duggan: "Machine annotation of traditional Irish dance music." PhD diss., Dublin Institute of Technology, 2009.

[7] K. Dunlay: "'Correctness' in Cape Breton fiddle music." *MacKinnon's Brook: Traditional Fiddle Music of Cape Breton, Volume 4*. Rounder Records 7040 Liner Notes, 2008.

[8] J. Duval: "Porteurs de pays à l'air libre: jeu et enjeux des pièces asymétriques dans la musique traditionnelle du Québec." PhD diss., Université de Montréal, 2013.

[9] E. Favreau: "La transmission de la musique traditionnelle par la radio." *Bulletin Mnémo* 2 (1) 1997. http://www.mnemo.qc.ca/html2/97(11)a.html (accessed 3 May 2015).

[10] Goertzen, C.: "Texas contest fiddling: What modern variation technique tells us." In *Routes & roots: Fiddle and dance studies from around the North Atlantic 4*, edited by I. Russell, and C. Goertzen, 98–111. Aberdeen: The Elphinstone Institute. 2012.

[11] C. Gore: *The Scottish fiddle music index: The 18th & 19th century printed collections*. Musselburgh, Scotland: Amaising. 1994.

[12] ———: "The Scottish music index." 2015. http://www.scottishmusicindex.org (accessed 3 May 2015).

[13] L. Hart, and G. Sandell: *Danse ce soir: Fiddle and accordion music of Québec*. Pacific, MO: Mel Bay. 2001.

[14] IREPI: "Yvon Mimeault, violoneux." *L'Inventaire des ressources ethnologiques du patrimoine immatériel* 2015. http://www.irepi.ulaval.ca/fiche-yvon-mimeault-86.html

[15] J. Keith: "The session." 2015. https://thesession.org (accessed 3 May 2015).

[16] M. Kelly: "Evaluation of melody similarity measures." MSc diss., Queen's University (Ontario), 2012.

[17] Kleinberg, J., and É. Tardos: "Dynamic programming." In *Algorithm design*, 251–335. Boston: Pearson/Addison-Wesley. 2006.

[18] A. Kuntz, and V. Pelliccioni: "Traditional tune archive." 2015. http://tunearch.org/wiki/TTA (accessed 3 May 2015).

[19] G. Labbé: *Musiciens traditionnels du Québec (1920–1993)*. Montréal: VLB. 1995.

[20] Library and Archives Canada: "The virtual gramophone." 2015. http://www.collectionscanada.gc.ca/gramophone/index-e.html (accessed 3 May 2015).

[21] M. Mongeau, and D. Sankoff: "Comparison of musical sequences." *Computers and the Humanities* 24161–75, 1990.

[22] M. Ó Súilleabháin: "Innovation and tradition in the music of Tommie Potts." PhD diss., Queen's University (Belfast), 1987.

[23] L Ornstein: "A life of music: History and repertoire of Louis Boudreault, traditional fiddler from Chicoutimi, Quebec." MA diss., Université Laval, 1985.

[24] R. Spottswood: *Ethnic music on records: A discography of ethnic recordings produced in the United States, 1893–1942. Vol. 1: Western Europe*. Urbana: University of Illinois. 1991.

[25] R. Thérien: *L'Histoire de l'enregistrement sonore au Québec et dans le monde 1878–1950*. Québec: Presses de l'Université Laval. 2003.

[26] P. Van Kranenburg: "A computational approach to content-based retrieval of folk song melodies." PhD diss., Utrecht University, 2010.