MODELING HARMONY WITH SKIP-GRAMS

David R. W. Sears **Andreas Arzt** **Harald Frostel**
Reinhard Sonnleitner **Gerhard Widmer**
Department of Computational Perception, Johannes Kepler University, Linz, Austria
david.sears@jku.at

ABSTRACT

String-based (or *viewpoint*) models of tonal harmony often struggle with data sparsity in pattern discovery and prediction tasks, particularly when modeling composite events like triads and seventh chords, since the number of distinct n-note combinations in polyphonic textures is potentially enormous. To address this problem, this study examines the efficacy of *skip-grams* in music research, an alternative viewpoint method developed in corpus linguistics and natural language processing that includes sub-sequences of n events (or n-grams) in a frequency distribution if their constituent members occur within a certain number of skips.

Using a corpus consisting of four datasets of Western classical music in symbolic form, we found that including skip-grams reduces data sparsity in n-gram distributions by (1) minimizing the proportion of n-grams with negligible counts, and (2) increasing the coverage of contiguous n-grams in a test corpus. What is more, skip-grams significantly outperformed contiguous n-grams in discovering conventional closing progressions (called *cadences*).

1. INTRODUCTION

Corpus studies employing string-based (or *viewpoint*) methods in music research often suffer from the *contiguity fallacy*—the assumption that note or chord events on the musical surface depend only on their immediate neighbors. For example, in symbolic music corpora, researchers often divide the corpus into contiguous sequences of n events (called n-grams) for the purposes of pattern discovery [4], classification [5], similarity estimation [16], and prediction [17]. And yet since much of the world's music is hierarchically organized such that certain events are more stable (or prominent) than others [1], *non-contiguous* events often serve as focal points in the sequence [11]. As a consequence, the contiguous n-gram method yields increasingly sparse distributions as n increases, resulting in the well-known *zero-frequency problem* [27], in which n-grams encountered in the test set do not appear in the training set. Perhaps worse, the most highly recurrent temporal patterns in tonal music—melodic formulæ, conventional chord progressions, etc.—are rarely included.

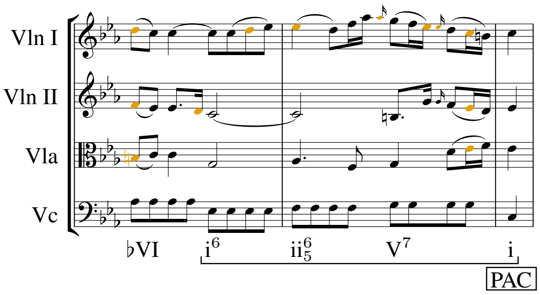

Figure 1: Haydn, String Quartet in C minor, Op. 17/4, i, mm. 6–8. Non-chord tones are shown with orange noteheads, and Roman numeral annotations appear below, with the chords of the perfect authentic cadence (PAC) progression embraced by a horizontal square bracket.

By way of example, consider the closing measures of the main theme from the first movement of Haydn's string quartet Op. 17, No. 4, shown in Figure 1. The passage culminates in a *perfect authentic cadence*, a syntactic closing formula that features a conventional chord progression (V–I) and a falling upper-voice melody ($\hat{2}$–$\hat{1}$). In the music theory classroom, students are taught to reduce this musical surface to a succession of chord symbols, such as the Roman numeral annotations shown below. Yet despite the ubiquity of this pattern throughout the history of Western tonal music, string-based methods generally fail to retrieve this sequence of chords due to the presence of intervening non-chord tones (shown in orange), a limitation one study has called the *interpolation problem* [3].

To discover the organizational principles underlying tonal harmony using data-driven methods, this study examines the efficacy of *skip-grams* in music research, an alternative viewpoint method developed in corpus linguistics and natural language processing that includes subsequences in an n-gram distribution if their constituent members occur within a certain number of skips. In language corpora, skip-grams have been shown to reduce data sparsity in n-gram distributions [13], discover multi-word expressions (or *collocations*) in pattern discovery tasks [22], and minimize model uncertainty in word prediction tasks [12].

Models for the discovery of harmonic progressions in polyphonic corpora typically exclude higher-order sequences (when $n > 2$) due to the sparsity of their distributions [18], so this paper examines the utility of skip-grams for 2-grams, 3-grams, and 4-grams. We begin in Section 2 by describing the *voice-leading type* (VLT), an optimally reduced chord typology that models every possible combination of note events in the dataset, but that reduces the number of distinct chord types based on music-theoretic principles. Following a formal definition of skip-grams in Section 3, Section 4 describes the datasets used in the present research and then presents the experimental evaluations, which consider whether skip-grams reduce data sparsity in n-gram distributions by (1) minimizing the proportion of rare n-grams (i.e., that feature negligible counts), and (2) covering more of the contiguous n-grams in a test corpus. We conclude by considering avenues for future research.

2. DATA-DRIVEN CHORD TYPOLOGIES

Corpus studies in music research often treat the *note* event as the unit of analysis, examining features like chromatic pitch [18], melodic interval [23], and chromatic scale degree [15]. Using computational methods to identify *composite* events like triads and seventh chords in complex polyphonic textures is considerably more complex, since the number of distinct n-note combinations associated with any of the above-mentioned features is enormous.

To derive chord progressions from symbolic corpora using data-driven methods, many music analysis software frameworks perform a *full expansion* of the symbolic encoding, which duplicates overlapping note events at every unique onset time.[1] Shown in Figure 2, expansion results in the identification of 23 unique onset times. Since expansion is less likely to under-partition more complex polyphony compared to other partitioning methods [4], we adopt this technique for the analyses that follow.

To reduce the vocabulary of potential chord types, previous studies have represented each chord according to the simultaneous relations between its note-event members (e.g., vertical intervals) [21], the sequential relations between its chord-event neighbors (e.g., melodic intervals) [4], or some combination of the two [19]. The skip-gram method can model any of these representation schemes, but for the purposes of this study, we have adopted the *voice-leading type* (VLT) representation developed in [19, 20], which produces an optimally reduced chord typology that still models every possible combination of note events in the dataset. The VLT scheme consists of an ordered tuple (S, I) for each chord in the sequence, where S is a set of up to three intervals above the bass in semitones modulo the octave, resulting in 13^3 (or 2197) possible combinations;[2] and I is the melodic interval (again modulo the octave) from the preceding bass note to the present one.

Because the VLT representation makes no distinction between chord tones and non-chord tones, the syntactic domain of voice-leading types is still very large. To reduce the domain to a more reasonable number, we have excluded pitch class repetitions in S (i.e., voice doublings), and we have allowed permutations. Following [19], the assumption here is that the precise location and repeated appearance of a given interval are inconsequential to the identity of the chord. By allowing permutations, the major triads $\langle 4, 7, 0\rangle$ and $\langle 7, 4, 0\rangle$ therefore reduce to $\langle 4, 7, \perp\rangle$. Similarly, by eliminating repetitions, the chords $\langle 4, 4, 10\rangle$ and $\langle 4, 10, 10\rangle$ reduce to $\langle 4, 10, \perp\rangle$. This procedure restricts the domain to 233 unique VLTs when $n = 1$ (i.e., when I is undefined). Figure 2 presents the VLT encoding for the PAC progression annotated in Figure 1, with the vertical interval classes S provided below each chord onset, and the melodic interval classes I inserted under horizontal angle brackets.

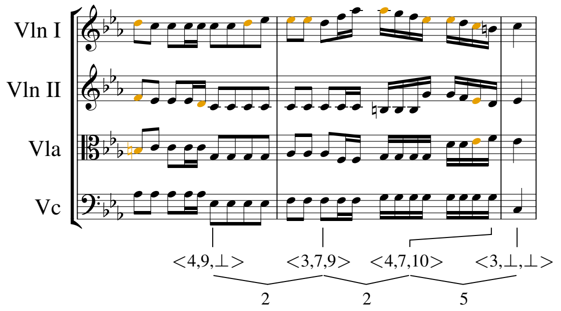

Figure 2: Full expansion of Op. 17/4, i, mm. 6–8. Non-chord tones are shown with orange noteheads, and the most representative chord onsets of the PAC progression are annotated with the VLT scheme.

3. DEFINING SKIP-GRAMS

In corpus linguistics, researchers often discover recurrent patterns by dividing the corpus into n-grams, and then determining the number of instances (or *tokens*) associated with each unique n-gram *type* in the corpus. N-grams consisting of one, two, or three events are often called *unigrams*, *bigrams*, and *trigrams*, respectively, while longer n-grams are typically represented by the value of n.

3.1 Contiguous N-grams

Each piece m consists of a contiguous sequence of VLTs, so let k represent the length of the sequence in each piece, and let C denote the total number of pieces in the corpus. The number of contiguous n-gram tokens in the corpus is

$$\sum_{m=1}^{C} k_m - n + 1 \tag{1}$$

This formula ensures that the total number of tokens is necessarily smaller than the total number of events in the sequence when $n > 1$.

3.2 Non-Contiguous N-grams

The most serious limitation of contiguous n-grams is that they offer no alternatives; every event depends only on its

[1] In *Humdrum*, this technique is called *ditto* [14], while *Music21* calls it *chordifying* [6].

[2] The value of each vertical interval is either undefined (denoted by ⊥), or represents one of twelve possible interval classes, where 0 denotes a perfect unison or octave, 7 denotes a perfect fifth, and so on.

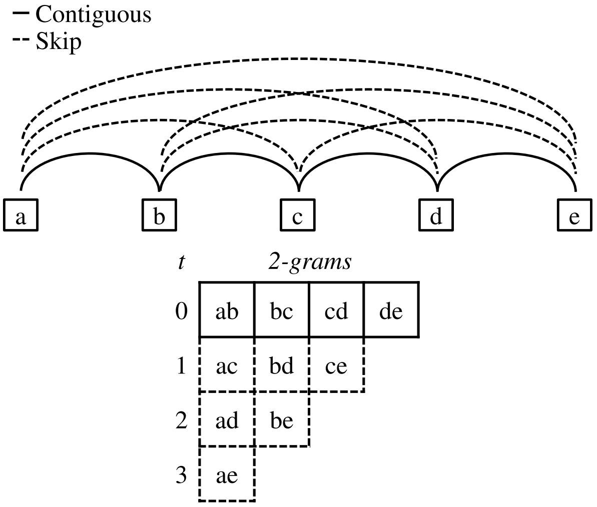

Figure 3: Top: A 5-event sequence, with arcs denoting all contiguous (solid) and non-contiguous (dashed) 2-gram tokens. Bottom: All 2-gram tokens, with t indicating the number of skips.

immediate neighbors. Without this limitation, the number of associations between events in the sequence necessarily explodes in combinatorial complexity as n and k increase.

The top plot in Figure 3 depicts the contiguous and non-contiguous 2-gram tokens for a 5-event sequence with solid and dashed arcs, respectively. According to (1), the number of contiguous 2-grams in a 5-event sequence is $k - n + 1$, or 4 tokens. If all possible non-contiguous relations are also included, the number of tokens is given by the combination equation:

$$\binom{k}{n} = \frac{k!}{n!(k-n)!} = \frac{k(k-1)(k-2)\dots(k-n+1)}{n!} \tag{2}$$

The notation $\binom{k}{n}$ denotes the number of possible combinations of n events from a sequence of k events. By including the non-contiguous associations, the number of 2-grams for a 5-event sequence increases to 10. As n and k increase, the number of patterns can very quickly become unwieldy: a 20-event sequence, for example, contains 190 possible 2-grams, 1140 3-grams, 4845 4-grams, and 15,504 5-grams.

3.2.1 Fixed-Skip N-grams

To overcome the combinatoric complexity of counting tokens in this way, researchers in natural language processing have limited the investigation to what we will call *fixed-skip n*-grams [13], which only include n-gram tokens if their constituent members occur within a fixed number of skips t. Shown in the bottom plot in Figure 3, ac and bd constitute 1-skip tokens (i.e., $t = 1$), while ad and be constitute 2-skip tokens. Thus, up to 7 tokens occur when $t = 1$, up to 9 occur when $t = 2$, and up to 10 occur when $t = 3$.

3.2.2 Variable-Skip N-grams

For natural language texts, the temporal structure of a sequence of linguistic utterances is not clearly defined. Yet for music corpora, temporal characteristics like onset time and duration play an essential role in the realization and reception of musical works. For example, the upper boundary under which listeners can group successive events into temporal sequences is around 2s [10]. Thus, as an alternative to the fixed-skip method, we also include *variable-skip n*-grams, which include n-gram tokens if the inter-onset interval(s) (IOI) between their constituent members occur within a specified upper boundary (e.g., 2s).

4. EXPERIMENTAL EVALUATIONS

This section describes the datasets in the present research and then examines whether the inclusion of skip-grams (1) minimizes the proportion of n-gram types with negligible counts, and (2) covers more of the contiguous n-gram tokens in a test corpus.

4.1 Datasets & Pre-Processing

Shown in Table 1, this study includes four datasets of Western classical music that feature symbolic representations of both the notated score (e.g., metric position, rhythmic duration, pitch, etc.) and a recorded expressive performance (e.g., onset time and duration in seconds, velocity, etc.). Altogether, the corpus totals over 20 hours of music.

The **Kodály/Haydn** dataset consists of 50 Haydn string quartet movements encoded in MIDI format [21]. The data were manually aligned at the downbeat level to recorded performances by the Kodály Quartet, and then the onset time for each chord event in the symbolic representation was estimated using linear interpolation.

The **Batik/Mozart** dataset consists of 13 complete Mozart piano sonatas encoded in MATCH format [24]. The data were aligned to performances by Roland Batik that were recorded on a Bösendorfer SE 290 computer-controlled piano, which is equipped with sensors on the keys and hammers to measure the timing and dynamics of each note [25].

The remaining two datasets were encoded in MusicXML format, and were also aligned to performances that were recorded on a Bösendorfer computer-controlled piano. The **Zeilinger/Beethoven** dataset consists of 9

Composer (Performer)	N_{pieces}	N_{chords}	$N_{\text{tokens}>3}$
Haydn (Kodály)	50	73,704	0
Mozart (Batik)	39	63,418	969
Beethoven (Zeilinger)	30	42,157	910
Chopin (Magaloff)	156	147,871	3666
Total	275	327,150	5545

Note. $N_{\text{tokens}>3}$ denotes n-gram tokens that initially consisted of more than three interval classes.

Table 1: Datasets and descriptive statistics for the corpus.

complete Beethoven piano sonatas performed by Clemens Zeilinger [8], while the **Magaloff/Chopin** dataset consists of 156 Chopin piano works that were performed by Nikita Magaloff [8, 9].

Performing a full expansion on all four datasets produced 327,150 unique onsets from which to derive chords. Unfortunately, some onsets presented more than three vertical interval classes, but since the VLT scheme only permits up to three interval classes S above the bass, it was necessary to replace these chords. Each onset containing more than three distinct vertical interval classes was replaced either with (1) the closest maximal subset estimated from the immediate surrounding context (i.e., ±5 chords); (2) the most common maximal subset estimated from the entire piece; or finally (3) the most common maximal subset estimated from all pieces in the corpus.

4.2 Reducing Sparsity

In natural language corpora, n-gram distributions of individual words ($n = 1$) and multi-word expressions ($n < 5$) demonstrate a power-law relationship between frequency and rank, with the most frequent (i.e., top-ranked) types accounting for the majority of the tokens in the distribution [26]. In music corpora, however, this relationship becomes increasingly linear as n increases due to the greater proportion of types featuring negligible counts. Such rare n-grams are thus more difficult to retrieve and model in discovery and prediction tasks, so this section examines whether the inclusion of skip-grams minimizes the proportion of rare n-grams in chord distributions.

4.2.1 Methods

Contiguous n-gram distributions were calculated from $n = 1$ to $n = 7$, along with 4-grams that include the following skip levels: *Fixed* – up to 1, 2, 3, or 4 skips; *Variable* – all possible skips occurring within a maximum IOI of .5, 1, 1.5, or 2s.

Skip	N_{types}	N_{tokens}
No Skip		
	135,331	326,034
Fixed – Skip boundary (#)		
1	850,222	2,604,972
2	2,364,840	8,780,643
3	4,765,289	20,786,976
4	8,207,123	40,548,000
Variable – IOI[a] *boundary* (*s*)		
0.5	2,213,148	10,150,852
1	12,498,736	90,278,381
1.5	31,591,468	306,289,766
2	59,147,107	718,717,231

[a] IOI denotes the maximum permitted inter-onset interval in seconds between adjacent members of each n-gram.

Table 2: Counts associated with 4-gram types and tokens using both fixed and variable skips.

4.2.2 Results

Table 2 presents the counts for 4-gram types and tokens with both fixed and variable skips. As expected, including skips of either type significantly increased the number of types and tokens. When skips were not included, the corpus produced over 300 thousand tokens, but this number increased to over 40 million tokens for skip-grams including up to 4 skips, or over 700 million tokens for skip-grams including all skips occurring within an IOI of 2s.

To visualize the increasing impact of data sparsity on the n-gram distribution as n increases, the top plot in Figure 4 presents the cumulative probability distributions for contiguous n-gram types from $n = 1$ to $n = 7$. Types appearing to the right of each marker feature only one token in the corpus. When n is small, the distributions loosely conform to the family of power laws used in linguistics to describe the frequency-of-occurrence of words in language corpora, where a small proportion of types account for most of the encountered tokens. When n increases, however, the proportion of types featuring negligible counts also increases, resulting in increasingly uniform distributions.

Shown in the bottom plot in Figure 4, the power-law relationship returns in the 4-gram distributions when skips are included. What is more, the proportion of types featuring negligible counts also decreases, thereby minimizing

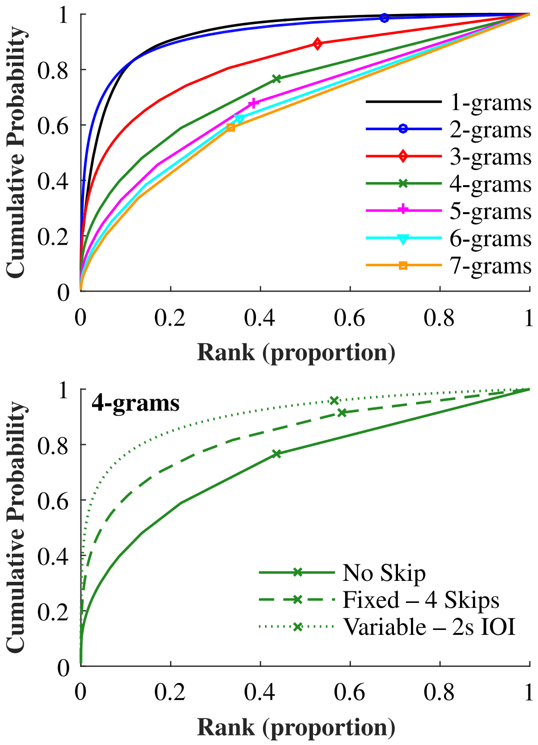

Figure 4: Cumulative probability distributions for (top) contiguous n-gram types, with types appearing to the right of each marker featuring only one token in the corpus; and (bottom) 4-gram types featuring no skips, up to four skips, or all skips occurring within an IOI of 2s.

the potential for data sparsity in the VLT distribution.

4.3 Increasing Coverage

This section examines whether the inclusion of skip-gram types during training covers more of the contiguous *n*-gram tokens in a test corpus.

4.3.1 Methods

2-gram, 3-gram, and 4-gram distributions were calculated for the following skip levels: *Fixed* – no skip, or up to 1, 2, 3, or 4 skips; *Variable* – no skip, or all possible skips occurring within an IOI of .5, 1, 1.5, or 2s. To evaluate skip-gram coverage, we employed 10-fold cross-validation stratified by composer [7], using the proportion of contiguous *n*-gram types in the test set that appeared in the training set as a measure of performance. To create folds containing the same number of compositions *and* chords, we computed the mean number of chords that should appear in each fold m, and then selected the fold indices for which each fold (1) contained an approximately equal number of compositions, and (2) contained a total number of chords that was $\pm 1\%$ of m.

4.3.2 Analysis

To examine the potential increase in coverage at each successive (fixed or variable) skip, we calculated a planned comparison statistic that does not assume equal variances, called the Welch t test.[3] The mean of each skip was compared to the mean of the previous skip using backward-difference coding (e.g., *Fixed*: 2 skips vs. 1 skip, 3 skips vs. 2 skips, etc.). To minimize the risk of committing a Type I error, each comparison was corrected with Bonferroni adjustment, which divides the significance criterion by the number of planned comparisons.

4.3.3 Results

Figure 5 displays line plots of the mean proportion of contiguous *n*-gram tokens from the test that appeared during training using either fixed or variable skips. Table 3 provides the mean coverage estimates and planned comparisons. For 2-grams, on average the contiguous types covered nearly 96% of the tokens in the test set. When skips were included, this estimate improved significantly to 98.3% of the tokens for up to two fixed skips, or up to 99.2% percent of the tokens for all skips occurring within an IOI of 1.5 s.

As n increased, the proportion of tokens that appeared during training using contiguous *n*-grams decreased substantially. For 3-grams, the contiguous types only covered 70.7% of the tokens on average. This estimate improved dramatically when either fixed or variable skips were included, however. For the fixed-skip factor, including up to four skips during training covered an additional 20% of the tokens during test, resulting in a mean coverage estimate of over 90%. In the variable-skip condition, this estimate further improved to 94.3% when all skips occurring within an IOI of 2s were included. Finally, for 4-grams, the contiguous types covered just 36.5% of the tokens, but this estimate improved to 71.1% in the fixed-skip condition, and to 82.4% in the variable-skip condition.

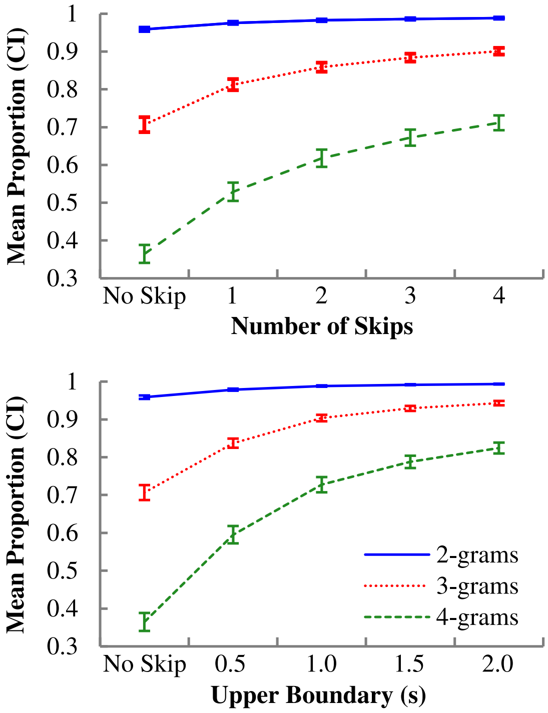

Figure 5: Line plots of the mean proportion of *n*-gram tokens from the test that were covered during training using either fixed (top) or variable (bottom) skips. Whiskers represent the 95% confidence interval (CI) around the mean.

5. SUMMARY AND CONCLUSION

To reduce data sparsity in *n*-gram distributions of tonal harmony, this study examined the efficacy of skip-grams, an alternative viewpoint method that includes sub-sequences in an *n*-gram distribution if their constituent members occur within a certain number of skips (*fixed*), or a specified temporal boundary (*variable*). To that end, we compiled four datasets of Western classical music that feature symbolic representations of the notated score. Our findings demonstrate that the inclusion of skip-grams reduces sparsity in higher-order *n*-gram distributions by (1) minimizing the proportion of n-grams with negligible counts, thus recovering the power-law relationship between frequency and rank when $n < 5$ that was previously lost in the corresponding contiguous distributions, and (2) increasing the coverage of the contiguous n-grams in a test set, thereby mitigating the severity of the zero-frequency problem.

In our view, this approach would directly benefit tasks

[3] In hypothesis testing, planned comparisons typically follow an omnibus statistic like the F ratio, which indicates whether the differences between the means of a given factor are significant. In this case, the Welch F test was significant for every model, so we forgo reporting those statistics here, and instead simply report the planned comparisons, which indicate whether coverage *increased* significantly as the number of skips (or the size of the temporal boundary) increased.

	2-grams			**3-grams**			**4-grams**		
Skip	M_{coverage}	t	p	M_{coverage}	t	p	M_{coverage}	t	p
No Skip									
	.959			.707			.365		
Fixed – Skip boundary (#)									
1	.976	7.144	<.001	.813	9.726	<.001	.529	10.963	<.001
2	**.983**	4.000	.003	.859	5.518	<.001	.618	6.023	<.001
3	.986	2.529	.085	.884	3.620	.008	.672	3.948	.003
4	.988	1.848	.327	**.901**	2.814	.046	**.711**	3.063	.027
Variable – IOI boundary (*s*)									
0.5	.979	8.439	<.001	.837	12.744	<.001	.595	15.795	<.001
1	.988	6.598	<.001	.904	10.132	<.001	.727	9.786	<.001
1.5	**.992**	3.647	.010	.929	5.313	<.001	.788	5.266	<.001
2	.993	2.311	.132	**.943**	3.564	.009	**.824**	3.808	.005

Table 3: Mean coverage estimates and planned comparisons for 2-gram, 3-gram, and 4-gram tokens using either fixed or variable skips.

related to pattern discovery and prediction, since recurrent temporal patterns rarely appear on the musical surface, thereby forcing n-gram models to either exclude higher-order n-grams (e.g., where $n > 2$) due to the sparsity of the distributions, or calculate *escape probabilities* to accommodate patterns that do not appear (contiguously) in the training set [2]. Consider, for example, the two four-chord cadential progressions in Table 4: the *semplice* cadence, which features a dominant-to-tonic progression in root position (e.g., I^6-ii^6-V^7-I); and the *composta* cadence, which also features a six-four suspension above the cadential dominant (e.g., ii^6-"I$^6_4$"-V^7-I). These cadences are ubiquitous in music of the classical style, and yet the VLT configurations representing these progressions rarely appear on the surface; the semplice cadence *never* appears contiguously, while the composta cadence is featured in just seven pieces. When skips are included, however, the two progressions appear in 32 and 77 of the 245 pieces in the corpus, respectively.

Skip	I^6-ii^6-V^7-I	ii^6-"I$^6_4$"-V^7-I
No Skip		
	0	7
Fixed – Skip boundary (#)		
1	3	16
2	10	36
3	13	50
4	15	63
Variable – IOI^a *boundary* (*s*)		
0.5	5	8
1	10	33
1.5	21	51
2	32	77

Note. VLT encodings for these progressions appear in the major and minor mode, and feature the pre-dominant and dominant harmonies both with and without the seventh (e.g., ii^6 and ii$^6_5$).

Table 4: Number of pieces containing semplice or composta four-chord progressions using both fixed and variable skips.

Due to the combinatoric complexity of the task, one limitation of the skip-gram method is that execution times become unfeasible beyond certain values of n and t. Nevertheless, if the organizational principles underlying hierarchical stimulus domains like natural language or polyphonic music reflect limitations of human auditory processing, it seems reasonable to impose similar restrictions on the sorts of contiguous and non-contiguous relations the skip-gram method should model. Given the restrictions imposed in this study, retrieving all 4-gram tokens from a sequence of 1,000 chords using commodity hardware produced runtimes of less than 100ms in the largest fixed-skip condition ($t = 4$ skips), and less than 3s in the largest variable-skip condition ($t = 2$s), proving skip-gram modeling is entirely attainable in a research setting.

Of course, counting all possible skip-grams in this way assumes no a priori knowledge about the sorts of non-contiguous relations analysts might hope to discover. For example, collocation extraction algorithms in the NLP community typically exclude infrequent n-grams, or use parts-of-speech tags to privilege syntactically meaningful utterances [22]. Music researchers could adopt similar methods by excluding (or weighting) each n-gram by the temporal proximity or periodicity of its members [21], or privileging patterns that appear in strong metric positions or feature changes of harmony. Together with the skip-gram method, these techniques could usher in a new suite of inductive, data-driven tools for the discovery of musical organization.

6. ACKNOWLEDGMENTS

This research is supported by the European Research Council (ERC) under the EUs Horizon 2020 Framework Programme (ERC Grant Agreement number 670035, project "Con Espressione").

7. REFERENCES

[1] J. J. Bharucha and C. L. Krumhansl. The representation of harmonic structure in music: Hierarchies of stability as a function of context. *Cognition*, 13:63–102, 1983.

[2] J. G. Cleary and I. H. Witten. Data compression using adaptive coding and partial string matching. *IEEE Transactions on Communications*, 32(4):396–402, 1984.

[3] T. Collins, A. Arzt, H. Frostel, and G. Widmer. Using geometric symbolic fingerprinting to discover distinctive patterns in polyphonic music corpora. In D. Meredith, editor, *Computational Music Analysis*, pages 445–474. Springer International Publishing, Cham, 2016.

[4] D. Conklin. Representation and discovery of vertical patterns in music. In C. Anagnostopoulou, M. Ferrand, and A. Smaill, editors, *Music and Artifical Intelligence: Lecture Notes in Artificial Intelligence 2445*, volume 2445, pages 32–42. Springer-Verlag, 2002.

[5] D. Conklin. Multiple viewpoint systems for music classification. *Journal of New Music Research*, 42(1):19–26, 2013.

[6] M. S. Cuthbert and C. Ariza. music21: A toolkit for computer-aided musicology and symbolic music data. In J. S. Downie and R. C. Veltkamp, editors, *Proc. 11th International Society for Music Information Retrieval (ISMIR)*, pages 637–642, 2010.

[7] T. G. Dietterich. Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation*, 10(7):1895–1923, 1998.

[8] S. Flossmann. *Expressive Performance Rendering with Probabilistic Models — Creating, Analyzing, and Using the Magaloff Corpus*. Phd thesis, Johannes Kepler University, Linz, Austria, 2010.

[9] S. Flossmann, W. Goebl, M. Grachten, B. Niedermayer, and G. Widmer. The Magaloff project: An interim report. *Journal of New Music Research*, 39(4):363–377, 2010.

[10] P. Fraisse. Rhythm and tempo. In D. Deutsch, editor, *The Psychology of Music*, pages 149–180. Academy Press, New York, 1982.

[11] R. O. Gjerdingen. "Historically informed" corpus studies. *Music Perception*, 31(3):192–204, 2014.

[12] J. T. Goodman. A bit of progress in language modeling. *Computer Speech & Language*, 15:404–434, 2001.

[13] D. Guthrie, B. Allison, W. Liu, L. Guthrie, and Y. Wilks. A closer look at skip-gram modelling. In *Proc. 5th International Conference on Language Resources and Evaluation (LREC-2006)*, pages 1222–1225. European Language Resources Association, 2006.

[14] D. Huron. *The Humdrum Toolkit: Software for Music Research*. Center for Computer Assisted Research in the Humanities, Stanford, CA, 1993.

[15] E. H. Margulis and A. P. Beatty. Musical style, psychoaesthetics, and prospects for entropy as an analytic tool. *Computer Music Journal*, 32(4):64–78, 2008.

[16] D. Müllensiefen and M. Pendzich. Court decisions on music plagiarism and the predictive value of similarity algorithms. *Musicæ Scientiæ*, Discussion Forum 4B:257–295, 2009.

[17] M. T. Pearce. *The Construction and Evaluation of Statistical Models of Melodic Structure in Music Perception and Composition*. Phd thesis, City University, London, 2005.

[18] M. T. Pearce and G. A. Wiggins. Improved methods for statistical modelling of monophonic music. *Journal of New Music Research*, 33(4):367–385, 2004.

[19] I. Quinn. Are pitch-class profiles really "key for key"? *Zeitschrift der Gesellschaft der Musiktheorie*, 7:151–163, 2010.

[20] I. Quinn and P. Mavromatis. Voice-leading prototypes and harmonic function in two chorale corpora. In C. Agon, E. Amiot, M. Andreatta, G. Assayag, J. Bresson, and J. Manderau, editors, *Mathematics and Computation in Music*, pages 230–240. Springer, Heidelberg, 2011.

[21] D. R. W. Sears. *The Classical Cadence as a Closing Schema: Learning, Memory, and Perception*. Phd thesis, McGill University, Montreal, Canada, 2016.

[22] F. Smadja. Retrieving collocations from text: Extract. *Computational Linguistics*, 19(1):143–177, 1993.

[23] P. G. Vos and J. M. Troost. Ascending and descending melodic intervals: Statistical findings and their perceptual relevance. *Music Perception*, 6(4):383–396, 1989.

[24] G. Widmer. Using AI and machine learning to study expressive music performance: Project survey and first report. *AI Communications*, 14(3):149–162, 2001.

[25] G. Widmer. Discovering simple rules in complex data: A meta-learning algorithm and some surprising musical discoveries. *Artificial Intelligence*, 146:129–148, 2003.

[26] J. Williams, P. R. Lessard, S. Desu, E. M. Clark, J. P. Bagrow, C. M. Danforth, and P. S. Dodds. Zipf's law holds for phrases, not words. *Scientific Reports*, 5(12209), 2015.

[27] I. H. Witten and T. C. Bell. The zero-frequency problem: Estimating the probabilities of novel events in adaptive text compression. *IEEE Transactions on Information Theory*, 37(4):1085–1094, 1991.