Darwin College Research Report
DCRR-004

Towards modelling harmonic movement in music: Analysing properties and dynamic aspects of pc set sequences in Bach's chorales

Martin Rohrmeier

16 May 2006

Darwin College
Cambridge University
United Kingdom CB3 9EU
www.dar.cam.ac.uk/dcrr

ISSN 1749-9194

Table of Contents

Another Appendix B including figures, diagrams and tables is included as an extra booklet.

1 Introduction: Tonality, cognition and probability

How musical elements are ordered and put into meaningful structures characterises one central aspect of the description of a musical system, or a particular musical style in music analysis. Accordingly, tonality (of a certain historical time) can be understood as a complex system of relationships between musical elements.

Temperley (2004) cites one paragraph by Meyer (1967):

> “Once a musical style has become part of the habitual responses of composers, performers, and practiced listeners it may be regarded as a complex system of probabilities [...] Out of such internalized probability systems arise the expectations — the tendencies — upon which musical meaning is built [...] The probability relationships embodied in a particular musical style together with the various modes of mental behavior involved in the perception and understanding of the materials of the style constitute the norms of the style.”
> (Meyer 1967: 8-9)

One can assent to Temperley's remark that “these words ring profoundly true”. They state a fundamental relationship between music style, music cognition and musical meaning. Furthermore, they suggest an interplay and a mutual dependency between expectancy, which governs musical understanding in the end, and the norms of a particular style. Musical style forms individual internalised musical knowledge, but, at the same time, is constantly renewed and carried on by musical knowledge of members of the society (compare Luhmann 1990; 2000). Moreover, musical knowledge and expectation, which is tightly linked with music emotionality (Meyer 1956), are framed by cultural and biological conditions (Cross 2003; 2004). Supposedly, internalised musical knowledge is not governed (only) by a consistent set of (taught) rules. It is rather developed by exposure and active interaction with a society's musical practices and incorporated through the acquisition of patterns, prototypes and behavioural regularities, such as dance, its context and social connotations, i.e., through processes of acculturation. Although this renders music and musical structures as inseparably linked to human biology, society, musical practices, and historical context, one essential part of the system may be described by analysing musical structure and its probabilitistic features. This description will have to be incomplete as it cannot incorporate various possibly covert influences, for instance the physiology of the human hand influencing piano style (Sudnow 2001; further compare similar studies by Blacking (1961) or Baily (1985)), but will be a description of partial, though nevertheless fundamental aspects related to musical style and its cognition.

Meyer's statement above, as well as his more concrete claim that “styles in music are basically complex systems of probability relationships in which the meaning of any term or series of terms depends upon its relationships with all other terms possible within the style system" (Meyer 1956: 54), appear (particularly in their invocation of Piston's (1941/1978) "table of usual chord progressions") to make direct appeal to statistical descriptions of musical styles. However, this statement may not be interpreted too literally, for, surely, it would be an oversimplified claim that one musical element determines its successor to a certain probability; even a context of several elements may not wholly determine the next element. Not only local, but also global factors and structural factors, such as musical form or motivic structure, influence the order and the freedom of choice of elements. Therefore, Meyer's statement will not have a simple translation into a model of tonality, which, like a Markovian approach, locally analyses chains of elements in probabilistic terms. An entire musical system of a certain (historical) society can be expected to have historically grown in a complex way, and to be characterised by a jungle of entangled influences and dependencies. Its rules may mix systematic and arbitrary aspects. The tonal system, like language, may be governed rather by vague rules/concepts, open boundaries, Wittgensteinian (1953) family relationships and Roschian (1978) prototypes.

Whereas a systematic, rule-based description of the musical system may bring with it a number of overt and covert problems, statistical approaches might yield important results. In this context, statistical results comprise global and rough relationships between elements once they are analysed as they appear empirically. Without disentangling dependencies in complex systems, simple insights about (overall) structural relationships in complex systems may be gained. Though these results are approximative in nature and may not involve deeper explications on structural causes, they may serve as viable approaches once they are embedded into an adequate framework of explanation. From this perspective, Meyer's statement can be understood as a rich metaphor concerning music cognition and musical style, which motivates the application of statistical methods in the description of musical systems yet minimises the amount of a priori knowledge/assumptions required another perspective which involve less a priori knowledge/assumptions about the musical system. This is particularly relevant when there is not much known about the network of influences and dependencies of a musical system, or for ethnomusicological research (e.g. Freeman & Merriam (1956), Lomax (1968)). Furthermore, it may turn out that simple probabilistic rules already describe a system or its cognition relatively well, whereas rule-based characterisations may be incomparably more complex. Therefore, statistical approaches appear also to be promising for investigating and describing tonality, particularly in order to explore and describe large scale regularities and overall organisation in tonality.

Statistical approaches to music have been proposed since the 1950s (Pinkerton 1956, Youngblood 1958; Kraehenbuehl (1958); Coons & Kraehenbuehl 1958; Cohen 1962; Hiller & Fuller 1967), and have seen a recent renaissance in approaches like Beran (2004), Beran & Mazzola (1999; 2001), Temperley (2004, 2002) and Zanette (2005 i.p.). However, whereas most research has focussed on melody and melodic sequences, not much has been done on vertical patterns in music, i.e., harmony.

2 Scope of this research

The intent of this project is twofold: the first part aims to address this gap with a (computational) case study on statistical aspects of vertical/harmonic musical structures in Bach's chorales. The chorales form a relatively homogeneous corpus which realises a characteristic historical instance of tonality. The vertical patterns are here modeled by pitch class set (after Forte 1973; henceforth *pc set*) progressions and the results from statistical analyses will be compared to traditional characterisations of harmony. Furthermore, from a perspective in the spirit of Meyer, this statistical approach sheds light on music cognition, as well. Consequently, in the second part of the project, the statistical results will be applied to investigate key implications of pc set sequences. This motivates the development of a simple though partial cognitive model of the dynamics of implied key structure in harmonic sequences.

The following two parts of this chapter will further discuss this outline in terms of related approaches. Chapter 3 will present concrete methodological considerations in respect of harmony and segmentation and a formalisation which provides necessary clarifications for the computational analysis. Chapter 4 will give a detailed description of the chorales database and the steps which have been taken in preparing the database, such as finding doublets and determining the key of each chorale. Chapter 5 will discuss the findings about pc set distributions and pc set sequence occurrences. Chapter 6 will present a method for key finding of a pc set sequence which is based on the results from chapter 5. Chapter 7 will present some reflections on dynamic aspects of key cognition and a simple model of contextual key implication. Chapter 8 will provide general discussion of the results.

2.1 Related approaches

Several studies are related to the aim of the first part of this study by the ways in which they approach descriptions of musical structure and tonality. Hiller & Fuller (1967) combine a structural analysis with an information theoretical analysis of Webern's Symphony Op. 21. Youngblood (1958) analysed tone frequencies and (first-order) pitch class transitions in songs by Schubert, Mendelssohn and Schumann, as well as in Gregorian chant in order to reveal statistical evidence describing stylistic differences. He found some minor differences between the three romantic composers, but an overall similar distribution of frequencies of scale degrees. Krumhansl (1990) integrates these distributions within a broader characterisation of tonality by her notion of key profiles, which hold for tonality from a statistical as well as a psychological perspective, as her probe-tone-experimental based key profiles correlate strongly to statistical distributions, like Youngblood's results. Temperley (2004) employs Bayesian methods to compute the probability of pitch class configurations with reference to a key, which he uses to describe the “tonalness” of musical passages in relation to its key. Zanette (2005) links a study on tone distributions to implications on musical meaning.

The system of harmony, in particular, may be described in manifold ways: as harmony comprises systematic relationships of vertical structures, its properties may be approached by rule-based approaches, or by sequence analyses of general statistical methods, for instance, depending on the particular focus of analytical interest. Most traditional theories of harmony (e.g. Grabner 1974 or Gauldin 1997) describe harmony from a constraint-based perspective, accounting for structure indirectly by a larger set of constraining rules. Temperley & Sleator (1999) develop, with reference to Lerdahl & Jackendoff (1983), a preference-rule approach to harmonic analysis. In terms of a statistical approximation to harmonic structure, Budge (1943) analyses chord frequencies in tonal pieces. Similarly, Piston's (1941/1978) compilation of frequencies of common harmony transitions can be interpreted as an early statistical approach towards harmonic progressions. Schmuckler (1989) carried out a complementary psychological experiment investigating chord transitions. The results of these studies will be considered below in the light of the findings of the present project. Furthermore, Eberlein (1994) manually performed a statistical study on chord progressions in a smaller corpus.

However, one has to be aware that harmony is not a comprehensive representation of tonality. Polth (2001a; 2001b) points out strikingly that harmonic analysis does not suffice for a satisfactory description of musical functionality. After Polth, tonality and functionality, which encompasses the way that musical elements interact meaningfully, are complex forms of structurally organised and mutually inter-dependent tone material. However, this does not rule out single-perspective descriptions such as key profiles, Bayesian statistics, or probabilistic characterisations of harmonic sequences. Polth's reminder can be related to factor-centred analyses employing the core idea of Conklin & Witten's (1995) multiple viewpoint approach. Viewpoints are modeled as descriptors of one or only few particular features of musical structure. Whereas single viewpoints are themselves incomplete descriptions of musical structure, combining them into a multiple viewpoint system appears to be a powerful approach to analysing musical structure.

Computational methods are employed from a perspective which aims to relate both computational/statistical and traditional music analytical methods. Following the spirit of Hiller & Fuller (1967), both may combine in order to gain deeper insights into the nature and structure of tonality and tonal organisation in musical pieces. From the perspective of computer-aided analysis, there are a number of related approaches. In an early pioneer work, Winograd (1968) programmed a system which was capable of harmonic analysis of hand-segmented music, based on generative grammars (Chomsky 1957). Maxwell (1992) realised another system for harmonic analysis using a complex rule system. Smoliar (1980) developed a tool for computer-aided Schenkerian analysis.

The majority of studies dealing with sequences of musical elements from a statistical perspective employ Markovian techniques to characterise probabilities of these sequences. Here, Markovian models are characterised by their assumption that the probability of one event depends (approximately) on a fixed local context of previous elements, but not on global factors. Markovian approaches to musical sequences are commonly used for automated music generation, style replication or harmonisation since the 1950s. Brooks et al. (1957) employ Markovian methods for hymn generation. Hiller & Isaacson (1958) carry out four experiments in computational composition, including monody, four-part first species counterpoint and experimental music. Ames (1989) gives a comprehensive overview about Markovian approaches. Moreover, Cope's (1991) composition system employs statistical/Markovian techniques. Bod (2002) models melodic segmentation with a Markovian approach. As mentioned above, Conklin & Witten (1995) developed a multiple viewpoint technique, which extends the Markov Chain method by using multiple analyses independently and combining their results. Reis (1999) extended Conklin &

Witten's approach by introducing psychologically founded constraints for the segmentation and storage of *n*-grams. Alamkam et al. (1999) use a Markov Decision Network to generate polyphonic music. Ponsford et al. (1999) used Markovian approaches for generation of harmonic movement. Pearce and Wiggins (2004) review the most important mathematical techniques for Markovian methods, with particular reference to monophonic music modelling. Extended Markovian techniques, such as hidden Markov models, have been applied in music transcription (Raphael (2002), Takeda et al. (2002), Xi Shao et al. (2004)) or style recognition and music retrieval (e.g. Chai & Vercoe (2001), Pickens et al. (2003), Jin & Jagadish (2002)). Raphael and Stoddard (2004) employ hidden Markov Models to model harmonic analysis. For present purposes and because of time constraints on this project, Markovian techniques will just be partially employed, only for a small aspect of the key implication model. But the the special potential of hidden Markov models is recognised as a priority for future work.

The body of Bach's chorales is a common corpus for computational approaches to music. Here, it has been chosen because it is a well-known corpus, which is easily available in computer-readable format and because the generally rather homophonic chorale structure can be processed without very complex segmentation problems. Most computational studies involving Bach's chorales address style replication or harmonisation. For instance, Allan (2002), Allan & Williams (2005) employ hidden Markov Models to choral harmonisation, Hanlon & Ledlie (2002) combine a hidden Markov Model with a constraint based system. Ebcioglu (1992) implemented a rule-based export system for harmonisation. Hild et al. (1992), Höldner (2004), or Meyerson (2001) train neural networks for harmonisation. Particularly relevant for this project is Conklin (2002), who computes 32 shortest significant patterns of voice leading (three harmonies) which he associates with particularly frequent harmonic standard patterns.

2.2 Modelling key cognition

The second part of the project aims to produce a simple model of the cognition of key structure. “Underlying all aspects of analysis as an activity is the fundamental point of contact between mind and musical sound, namely musical perception” (Bent & Pople 2001). Most generally, music analysis, as well as music cognition, involves an act of interpretation which correlates the surface of a piece to an underlying (cognitive) structure (Cross 1998). The surface of a piece may be the raw acoustical or score-based information, whereas the structure as which it is interpreted may be a network of interrelations, for instance tonal dependencies from a Schenkerian or Riemannian perspective, or from that of Lerdahl & Jackendoff (1983) or Narmour (1989; 1992). Thus an act of analysis can be formally understood as a (possibly very complex) mapping between musical surface and a certain structure. From a cognitive perspective, it is not clear which components a structural interpretation of the musical surface would be composed of and if there are independent basic musical structures. Temperley (2001) discusses metrical, phrase, contrapuntal, pitch, harmonic and key structure – and it appears parsimonious to assume that key structure might be at least a basic component of music cognition, due to the simple fact the key structure is indeed an omnipresent (and defining) foundation of tonal music.

In a computational approach to music analysis, the difference between structure and surface has to be respected because its implications differ according to whether surface or structure data are analysed (Temperley (2004) stresses this difference, as well). Accordingly, the model of the dynamics of key structure represents the interpretation involved as a mapping between representations of the musical surface in terms of sequences of pc sets and the key structure in terms of keys assigned to those sequences.

3 Practical considerations and formalisation

In practical terms, the outlined analysis of statistical properties of vertical musical patterns outlined here can be realised in manifold ways, depending on the way these patterns are parsed/represented and the method of segmentation. The following sections will discuss considerations on this and present a mathematical formalisation of the performed method.

3.1 Approaching harmony

Vertical musical patterns, or harmony, can be represented and approached in various ways, such as in terms of abstract harmonic structure, actual vertically concurrent notes or intervals, or pc sets. Essentially, their differences consist in whether the focus is rather centred on musical surface or a structural interpretation of it.

Harmony is clearly a structural feature. It is an abstraction/reduction of the voice material which involves neglecting absolute pitch information in favour of (chromatic) pitch class information, considering different arrangements of the same pitch class material as identical chords in different inversions and involves identifying the root of the chord as referent. In a broader sense, harmonic analysis involves segmenting polyphonic texture into appropriate harmonic units and differentiating between harmonic and non-harmonic notes.

These aspects render harmony a complex task for computational analysis. For instance, identifying harmonic notes in cases where consonant notes exist as consecutives in the immediate context is a difficult task which requires a complex set of rules (e.g., the case of a suspended E and G underpinned by a movement from B to C). Furthermore, identifying the root is not straightforward as there are chords with identical surface but different harmonic structure, depending on the context. As a simple example, the same chord G-C-E can have two interpretations as second inversion of the C major triad, or a 6-4 suspension of a G major triad in a dominantic context (V^{6}_{4}) in C major. Similarly, incomplete chords such as A-C or C-G raise a problem in identifying their structural affiliation between binary possibilities (A-C-E/C-E-G; C-E-G/C-E*b*-G) reliably, although the context mostly clarifies their structure. Processing of these cases would involve a number of extremely complex and very task/style-specific contextual rules. Moreover, it is far from clear whether a greater proportion of harmonic mappings of musical surface is not intrinsically ambivalent.

Although this renders a computational analysis of harmony difficult, but not impossible (see Winograd (1968), Maxwell (1992), Pardo & Birmingham (2002), Barthélemy & Bonardi (2001)), it would nevertheless incorporate a rather large corpus of rule-based musical knowledge. But this is incongruent with the perspective of this study, which aims to connect bare statistical properties with cognitive aspects of basic musical structures. Therefore, a representation of vertical structure will be adopted which is closer to the musical surface and avoids taking up too much a priori music theoretical baggage. Furthermore, a segmentation method will be proposed which performs a reduction to pc sets by a metrical and dissonance based heuristic, which will be taken to be an analytical alternative in respect of some aspects of harmonic structures.

Conklin (2002) models vertical progressions as relative vectors of (intervallic) voice progressions. Similarly, vertical structures might be modeled by the actual set of concurrent notes, or by voice-leading excerpts from the outer voices. These approaches represent vertical patterns as very close to the surface and have their application once the actual shape of vertical surface is relevant to the analytical perspective (for instance, for a Schenkerian analysis). However, different inversions of the same pitch class material may not directly be considered as related, which will not support an analysis of similar patterns.

Therefore, vertical patterns will be represented as pc sets, here, following Forte (1973). A pc set representation reduces a chord to an unordered set of chromatic (enharmonically equivalent) pitch classes, disregarding absolute pitch and multiple repetitions of pitch classes. Hence, different arrangements or inversions of the same pitch material will be regarded as the same pc set. This representation is still relatively close to the musical surface without being a (theoretically loaded harmonic) interpretation. Furthermore, pc set sequences may be also interpreted as a viewpoint in the sense of Conklin & Witten (1995), which can be combined with other viewpoints which include bass, dissonance, or other contextual information – which may yield an interesting structure to be compared with harmonic analysis.

3.1.1 Segmentation

The second preliminary aspect of the computational analysis involves the segmentation of the voice-based polyphonic data into units of pc sets. One can think of several ways of segmenting a chorale but depending on the method of segmentation, the result obtained will be influenced.

A first possibility is a '*maximal*' segmentation which performs vertical cuts of the piece at each change of the smallest common time unit (fig. 3.1d). This way, a cut includes notes that are held across the smallest time division. Although this method ensures that all vertical patterns which can arise will be covered, it reduplicates held chords into larger chunks of a single repeated pc set. This causes the problem that the length of the sequences to be processed will be unnecessarily drastically increased.

A modified version of the 'maximal' method is to segment only at those time positions where at least one voice/note event changes (fig. 3.1a,e). This way, meaningless pc set repetitions are avoided and repetitions of a pc set indeed denote a change of voicing of the same pc set. However, this method[1] ('*dense segmentation*') treats all note events equally and there is no distinction between a semiquaver passing event and a crotchet chord. Thus, this method is useful to investigate the range and transitions of the whole set of pc sets occurring in a piece, but is inadequate to reveal the similarity of patterns which just differ slightly, are elaborations of a simpler structure or are rhythmical variations of each other[2].

When a segmentation is required which is more abstract and closer to a cognitive structure,

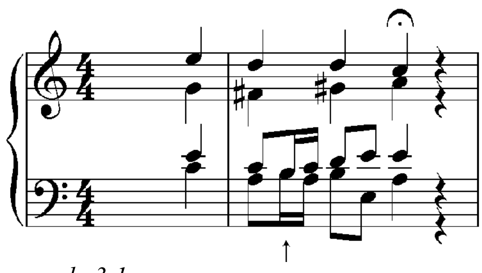
1 In Huron's *HUMDRUM,* a comparable processing method termed "full expansion" is implemented (Huron 1999).
2 The fact that this analysis ignores rhythmical aspects of pc set patterns may yield implications as to relations between otherwise unrelated sequences, as well.

methods appear appropriate which use metre or consonance as a cognitive cue. However, as a reduction to harmonically significant pc sets is too complex for present purposes, two approaches will be described which may serve as approximations. In the first, only those pc sets are considered significant which occur on stronger metrical positions (henceforth “metrical segmentation”, fig. 3.1b), omitting weaker quavers and semiquavers, are considered significant. For instance, Jørgensen & Madsen (2004) use this method; Ponsford et al. (1999) segment on quaver level. Metrical segmentation provides an uncomplicated way of 'sampling', but on the other hand, is not unproblematic. For instance, it will include stressed dissonances or passing notes and not their resolutions. A quaver resolution of G-C-D into G-B-D will be treated as G-C-D. Indeed, one might argue that this way of treating vertical structure might reveal interesting patterns as well; furthermore, it is unclear what a cognitively-informed treatment of such cases would be like. But nevertheless, this method appears to be overly simplistic and not ideal for a number of harmonic investigations.

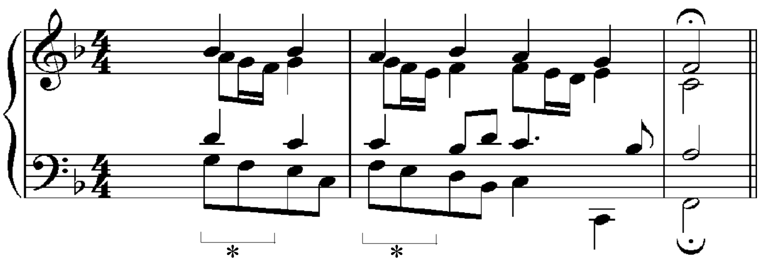
The second method aims to address these shortcomings in order to better approximate to harmonic features of the musical surface. It employs a selection process which chooses one harmonically representative pc set from all pc sets occurring within each crotchet beat. Although the practice of Bach's chorales shows that there might be no simple rule-based system which correctly identifies transitory dissonances, neighbour notes and suspensions and their resolutions correctly, the following simple rule appears to deal as a viable first approximation:

> (R1) If the first chords of a set is dissonant, the least dissonant chord of the set will be preferred. If the first chord is consonant or a dominant seventh chord, it will be preferred.

This may yield a better approximation to an adequate reduction than a simple *metrical segmentation.* Nevertheless, there are various cases where problematic results will be produced. Again, the case of a V^{6}_{4} chord which resolves into a V^{5}_{3} on the quaver level cannot be easily and reliably distinguished from an instance of a $I-V$ progression without reference to an embedding context. Furthermore, there are vertical passing phenomena which do not having anything to do with any harmonically relevant structure (ex.3.1). Similarly, example 3.2 shows the even more problematic (though rare) case of passing phenomena where the actual harmony is not even present as one concurrent simple vertical structure. Apparently, these cases, which stem from the underlying polyphonic structure and have a number of parallels in Bach's chorales, cannot be treated without a complex rule system[3] (Hoffman & Birmingham (2000) enter their harmonic segmentation manually). Furthermore, both reductive segmentation methods share the assumption of a steady harmonic rhythm on a crotchet level. This is already a rough-and-ready assumption with many exceptions, even though it underlies a number of computational approaches to Bach's chorales (e.g. Hild et al. (1992), Jørgensen & Madsen (2002)).



example 3.1



example 3.2

Nevertheless, the segmentation method employing rule R1 has been realised as *"harmony approximation"* (fig. 3.1c), as it appears to be a useful approximation notwithstanding the problems above. The notion of a correlation between stronger metrical position and harmonic relevance is an intuitively acceptable principle (for instance following Lerdahl & Jackendoff (1983) or Temperley & Sleator (1999)). Practically, the notion of "dissonance" has been modeled by a simple score system for pc sets. For computing the score, the (non symmetry-invariant) normal form and its interval vector are used (after Forte 1973).The score results as the sum of the occurrences of each interval multiplied by -4 for minor seconds, -1 for major seconds, -1 for tritones and 0 otherwise. The special case of an augmented triad is given a score of three. Furthermore, as triads and dominant sevenths appear to be particularly preferable, especially preferable to 0 scoring incomplete triads, triads are assigned a value of 2 and dominant sevenths (with and without fifth) a value of 1. Altogether, this realises a preference hierarchy of chords which is shown in table 3.1.

3.2 Mathematical formalism

In order to clarify their computational implementation, the methods described above have been mathematically formalised, mainly following and linking to the formalism of Pearce & Wiggins (2004).

For the present purpose, the (computational) analysis of the corpus will be performed in two steps: first, a segmentation of the representation of the musical surface of a piece into segments of analytical interests and, second, a transformation of the surface of these segments into distinct

3 This is not impossible, as Maxwell (1992) developed such a rule system which appears to function.

objects of interest from a finite or infinite alphabet ξ of symbols. The resulting codified piece can be analysed, for instance using techniques of *n*-gram (sequences of length *n*) analysis. Furthermore, an interpretation of a sequence of such objects into another structural representation (defined by another alphabet Υ), will be described as a mapping between sequences of these two domains/alphabets.

Due to the constraints of the MIDI format, a piece is represented as a sequences of note events n_i which are given as vectors of pitch (an integer representating the MIDI pitch), onset and duration time (in beats): $n_i \in Z$ with $Z = \mathbb{Q} \times \mathbb{N} \times \mathbb{Q}$, and $n_i = \langle o_i, p_i, d_i \rangle$. A piece, consisting of a sequence of note events, is characterised as $\langle n_i \rangle_i \in Z^*$ where Z^* denotes the set of all sequences of members of *Z*, including the empty sequence ε.

Describing the process of mapping a piece into a sequence of units to be analysed requires segmentation. A segment of the piece will be understood as a selection of note events, a segmentation as a set or sequence of segments. Note events may occur in several segments. Hence a segmentation can be defined as a sequence of sets of indices of the selected note events in each segment.

Accordingly, a segment will be characterised as a subset *k* of the set *L* of all indices: $k \in P(L)$ where *P* characterises the power set of *L*. A segmentation is a set of segments with an index set M: $S = \{k_i\}_{i \in M}$. Particularly, this reveals the following characterisation of the method of *dense segmentation*: $S = \{k(o_i)\}_{i \in L}$, $k(t) = \{i | o_i \leq t \wedge o_i + d_i > t\}$ [4]

In the case of *metrical segmentation*, only onset times at metrical beat onsets (which are integer values on beat level) stresses are selected: $S_2 = \{k(t)\}_{t \in N \cap [\min(o_i);\max(o_i)]}$.

In the case of the *harmonic approximation*, for each segment of 1 beat, the pc set is taken which scores best for the dissonance function *diss*, described above.

$$S_3 = \{h(t)\}_{t \in N \cap [min(o_i);max(o_i)]}, \quad h(t) = \underset{k(j) \text{ for } j \in [t;t+1)}{arg\,max} \left(diss\left(k(j)\right)\right)$$

For any segmentation *S*, the selected note events are: $\left\{\{n_j\}_{j \in k_i}\right\}_{i \in M} = \left\{\{\langle o_j, p_j, d_j \rangle\}_{j \in k_i}\right\}_{i \in M}$

4 Here, the fact that mathematical sets do not include double elements is used.

Furthermore, for the analysis here, both segmentation methods are chosen so that each segment is represented by a single onset $o(k_i)$ and a single duration $d(k_i) = \min\{d_i | o_i = o(k_i)\}$. Therefore, each segmentation can be described as $\{o(k_i), \{p_j\}_{j \in k_i}, d(k_i)\}_{i \in M}$.

This makes it possible to characterise each segment as one pc set, and the entire segmentation as a sequence of pc set events which are like note events characterised by onset and duration:
$\left\langle o(k_i), \tau\left(\{p_j\}_{j \in k_i}\right), d(k_i) \right\rangle_{i \in M}$ with $\tau : \mathbb{N}^* \to \xi, \{p_j\}_{j \in k_i} \mapsto \{p_j \ mod\, 12\}_{j \in k_i}$
Here, τ is the projection which transforms a set of pitches into a symbol from the alphabet ξ of pitch class sets.

Having segmented the piece, the structure can be analysed as sequences e_i, respectively *n*-grams of symbols from the alphabet ξ. Moreover, a sequence of events $e_i, ..., e_j$ for integer indices $i \leq j \in \mathbb{N}^+$ (a (*j*-*i*+1)-gram) will be denoted by $e_i^j \in \xi^*$. The number of different *n*-grams is denoted by $E(n) \in \mathbb{N}$. Further, the frequency of one pc set sequence in the database will be denoted by $c(e_i^j)$, and its probability by

$$p(e_i^j) = \frac{c(e_i^j)}{\sum_{e \in E(j-i+1)} c(e)} \qquad (1)$$

in terms of its relative frequency. The probability of an event occurring immediately after a certain context will be described by the *maximum likelihood* estimation (Pearce & Wiggins 2004):

$$p\left(e_i \middle| e_{(i-n)+1}^{i-1}\right) = \frac{c\left(e_i \middle| e_{(i-n)+1}^{i-1}\right)}{\sum_{e \in \xi} c\left(e \middle| e_{(i-n)+1}^{i-1}\right)} \qquad (2)$$

Pearce & Wiggins 2004 note that this method is rather simple and will become problematic in contexts of sparse data. However, it will serve for a simple application which does not aim to produce sequences, but just analyses the corpus of pieces because unknown contexts cannot occur in this case.
In this context the Markov assumption that the probability of the next event depends only on the previous *n*-1 ($n \in \mathbb{N}^+$) events can be formalised as:

$$p\left(e_i \middle| e_1^{i-1}\right) \approx p\left(e_i \middle| e_{(i-n)+1}^{i-1}\right)$$

The scheme of an analytical act presented above essentially involves an *interpretation* which maps musical surface to its underlying structure. For the present purpose of modelling key structure, it appears sufficient to model the key structure as a sequence of symbols from another alphabet Υ. Then, an interpretation of a sequence of surface symbols is defined as a function $\varphi : \xi^* \rightarrow \Upsilon^*$, which maps sequences of symbols from the alphabet ξ onto symbol sequences from the alphabet Υ. For this purpose, Υ denotes the set of 12 possible keys and each pc set symbol is mapped to a key symbol which characterises the entire sequence until that pc set. Accordingly, Υ can be defined as:

$\Upsilon = \{C, C\#, D, E^b, E, F, F\#, G, A^b, A, B^b, B, c, c\#, d, e^b, f, f\#, g, g\#, a, b^b, b\}$

4 Technology, Data and Preprocessing

For the present purpose, it appeared most reasonable to use MATLAB and the MIDI Toolbox, developed by Tuomas Eerola, http://www.jyu.fi/musica/miditoolbox/) as development platform. MATLAB provides a high level platform-independent programming framework and a variety of ready-to-use statistical and mathematical functionality. The MIDI Toolbox includes relevant I/O as well as basic preprocessing functionality, which considerably reduced the number of basic functions to be developed. However, one major disadvantage is MATLAB's slowness (a disadvantage which, though less severe, a Java based platform-independent implementation would have to face as well).

4.1 The MIDI format

The musical data is given in form of MIDI files. The MIDI standard (= Musical Instrument Digital Interface) provides a protocol and formal language for electronic musical instruments, which represents information on note events such as the pitch, onset, duration, as well as instrument (represented by an instrument number in a standardised list of MIDI instruments) and attack velocity (loudness), which is structured by several channels. The MIDI format is far from being a close representation of musical surface, or the musical score. (For the chorales, the MIDI information dealt with is close to the raw musical score, but it is not clear if it is representative for the cognition of chorale music as there may be peculiar additional acoustic/cognitive features) For the present purposes, the representation particularly lacks of information on fermatas – therefore, aspects of phrasing could not be included into the research, and segmentation algorithms such as Eerola (2003) turned out to be relatively unreliable for the chorales. Altogether, the available information for this study comprises key signature, metre (both Bach's original signatures), note onset and duration, given in beats. In the case of metre, another weakness of the MIDI format is revealed as the format only allows one specified metre, but there are chorales with changes of metre (B203[5]&204, B350). Generally, further description and discussion of the MIDI standard can be found in Selfridge-Field (1997). However, a future study may use (manually) annotated material in another musical format (such as the Humdrum/Kern format (see Huron 1997)).

4.2 Preparing the Database

The chorales were taken from the online database JSBChorales.net as MIDI - files

5 Henceforth, chorales will be identified by their number according to the Breitkopf edition.

(http://www.jsbchorales.net/sets.html) in the version from April 14, 2005[6]. It provides a rather large set of chorale material in comparison to the chorale set available in the Humdrum format http://dactyl.som.ohio-state.edu/Music824/databases.index.html. Nevertheless, a number of files needed preprocessing and minor errors in two files were found. Another randomly chosen sample of chorales was manually checked for note errors and did not contain note mistakes.

The downloaded set contained 521 midi files. It comprises the chorales from the Riemenschneider and the Kalmus edition. Several chorales with larger instrumental parts are found in a full and in a reduced version, which leaves out instrumental parts. A number of chorales contain one extra continuo part. Moreover, the set includes chorales from Cantatas, which are not included in the Riemenschneider or Kalmus editions (often with larger instrumental parts). Finally, it contains 42 chorales for organ. Altogether, there is 1 file with one part (midi channel), 10 files with 3, 381 files with 4, 69 files with 5,12 files with 6,15 files with 7, 32 files with 8, and one file with 9 parts. Files with fewer than four parts were excluded, and from files with more than 4 parts only the four part chorale set was taken. (It is assumed that in all files the four choral parts use midi channels 1-4, which was double checked in a number of sample files.) One chorale (B350) had to be excluded because it was a 5-part chorale. In a small number of cases, note durations which overlapped with the next note had been shortened.

For the present purpose only chorales which are found in either Riemenschneider or Kalmus editions were selected. However, the database (and this subset) contains a number of doublets (with and without transposition), which had to be excluded (using the detection algorithm described below). In all, the actual corpus which is analysed contains 386 pieces.

4.3 Finding doublets

The database contains doublets, some in the form of identical files, some as transpositions, some chorales which turn out to be different as soon as the instrumental parts are left out, and some as virtual doublets which just differ in a very small number of notes. As doublets (especially in terms of pc set progressions) strongly affect some of the analyses below, they had to be identified and excluded.

However, a brute-force exhaustive comparison between all pairs of files in all possible

6 The database is online (at least) since 1996 and has been checked and updated by numerous contributors. The history of corrections can be found at http://www.jsbchorales.net/correc.html.

key/transposition possibilities would have been too time consuming. Therefore, a method was developed and implemented to represent the chorales (their pc set progressions) independently of their key in order to make a simple sequence comparison possible. Accordingly, a sequence of absolute pc sets has to be represented relatively by applying an adequate homomorphism. The relative difference between two pc sets can be expressed effectively by the bitwise logical exclusive-or (XOR) operator applied to their binary representations. Then, a pc set sequence is represented by its initial pc set and the sequence of bitwise relative differences. In order to enable comparison these representations have to be normalised. This is achieved by transposing the initial pc set to the standard form (without inversion) after Forte (1973), which represents its 'kind' of pitch class set. Subsequently; the entire relative sequence is transposed by the interval necessary to put the initial set into standard form.

This method is not invariant in respect of key relation, i.e. it is not suitable for finding the key-independent scale degree representation of the file (this is described below), but *is* invariant in respect of the equivalence relation. Thus, the sequences can be described in a relative form and the transformed database can be searched for doublets employing the standard linear method. Additionally, another heuristic marked pieces as virtual doublets if their first 30 pc sets had been identical. Altogether, 61 doublets were found with the method above. This result was confirmed by an alternative method which is mentioned in chapter 5. Moreover, there are several pieces which are fairly similar, but differ (though often just slightly) by several pc sets . They have not been excluded from the dataset. (see also chapter 5)

4.4 Normalising

In order to compare properties of pc set sequences, the sequences have to be analysed relative to the key of the piece they are found in. Indeed, the chains have to be analysed with reference to the piece because otherwise, a relative (possibly profile-based) transposition algorithm would cut down the number of functionally different, but transpositionally identical chains (such as G-C-d in F-major, C-major or d-minor) enormously, which particularly affects results for shorter chains significantly (!).

Most computational approaches to Bach's chorales take this step of normalising in reference to key, but very few take into account the fact that the notion of key is not unproblematic for Bach chorales. Consideration must be given to the extent to which Bach's chorales fit into the modern major-minor key system and how (partial) modality and the number of modal chorales should be treated. Although in late Baroque, the use of modes (Kirchentöne) may have nearly disappeared (in respect of genres in *stylus luxurians* or *theatralis*), chorale-based music is still strongly connected to traditional chorales employing modal notation and melody which do not fit into the major-minor dualism (Daniel 2000). Whereas Ionian and Aeolian chorales transfer more or less easily into major and minor, Dorian, Mixolydian and Phrygian chorales raise problems when one attempts to sort them into the tonal system. However, it seems inadequate for an approach to neglect these modal chorales in a study of Bach's tonality (as do Knipphals/Möller 1995), as they occur frequently and constitute an essential component of Bach's chorale style (Burns 1995). Nevertheless, a single tonal centre has to be identified in respect of of each of these modal chorales. This can be undertaken in two particular ways.

Firstly, the *finalis* could be generally identified as the tonal centre. The main consequence of this is that Phrygian or Mixolydian chorales defined in this way tend not to be interpretable as tonally closed (for instance, they may end with an imperfect cadence). This criterion, on the other hand, would imply that their final cadence should exist in the system as a representation of a perfect, not an imperfect cadence. In theory, it is unclear which way Phrygian final cadences have to be interpreted, and there are examples in which the Phrygian *finalis* may be understood as the tonic. Daniel (2000:19) demonstrates that the literature is undecided on this issue, and finally finds that both interpretations can be grounded by equally legitimate reasons. However, for present purposes, it seems more appropriate to refer each example to a related major or minor key, because separating the chorales by modes or *finales* may fragment the data into rather sparse groups, and the statistical results (particularly on phrase endings) would probably be diversified and less significant. Secondly, associating each chorale to major or minor may be an adequate assumption in respect of music cognition. Therefore, the process of normalisation can be addressed as a modified standard key-finding problem. However, by so doing, modal chorales are nevertheless not reduced/subsumed into a major-minor system. They are transposed to C, but their idiosyncratic harmonic features will be reflected in the statistical results.

Keyfinding turned out to be a more time-consuming problem. The Krumhansl algorithm (Krumhansl 1990) did not yield useful results in a large number of cases, which confirms known weaknesses of the algorithm (see Temperley 2001).[7] The problems inhere in the problematic values embodied in the key profiles it presents, and the sometimes idiosyncratic and modulating character

7 Retrospectively, Krumhansl's algorithm classified 79.53%, Temperley's (2001) algorithm 20.47% of the chorales correctly.

of Bach's chorales (even on small segments at beginnings or endings). The Longuet-Higgins/Steedman (1971) and Holtzmann (1977) algorithms could not be used here as they are based on monophonic input and might not have yielded any reliable improvement. Therefore a specialised rule-based approach has been developed and applied. Taking into account the fact that the last chord of a chorale is the tonic in most of the cases (and a dominant in somewhat fewer cases) reduces the number of possible choices. In the first place, if the final chord is minor, it is assumed to be the tonic and the algorithm run ended, otherwise, it leaves three or four choices for the potential key.[8] Secondly, the key signature was taken from the information in the midi file. It was assumed (and was double checked with numerous samples) that the key signature information in the database was correct (only one exception was found). The key signature specifies a scale which leaves five possibilities for the actual tonic, which, for the instance of the empty key signature, is c or a for Ionian and Aeolian cases, d and g for the Dorian and Mixolydian cases, and the very rare possibility of e. Matching the induced key possibilities from the final chord and the key signature leaves no more than two possibilities for each case, which can be decided by applying a simple heuristic. Table 4.1 shows the range of possibilities for the empty key signature and all possibilities of final chords. The four ambiguous cases can be decided by simple preference rules, assuming that Mixolydian is unlikely to end on its dominant (case d,G), that the case of e minor, given an empty key signature is very rare (case e,a), that Ionian will not end on its dominant (G,C)[9] and that Dorian does not end on its dominant (a,d) (see Table 4.1). These rules yield a key estimation which turned out to be appropriate for most of the pieces, which renders the rules a reliable heuristic. Finally, Daniel (2000:18n.) discusses two chorales with have irregular key signatures. Both have been classified manually.

8 Employing the first chord of the piece as key-relevant turned out to yield too many exceptional cases to give reliable information on the piece's key.

9 This case, which distinguishes Ionian and Mixolydian modes may indeed be debatable for some chorales, for Mixolydian can be tonally very ambiguous. But altogether, it appeared to be a viable method for these cases.

Key rule 1: (final chord is minor)							
Final chord	***c***	***d***	***e***	***f***	***g***	***a***	***b***
Assigned key	***c***	***d***	***e***	***f***	***g***	***a***	***b***
Key rule 2: (final chord is major)							
Final chord	***C***	***D***	***E***	***F***	***G***	***A***	***B***
Possible finalis	c,f	d,g	e,a	f,b^b	g,c	a,d	b,e
Possible finales, given the key signature	(e) a d G C						
Remaining key possibilities	C	d,G	(e),a	-	G,C	a,d	e
Key heuristic	✓	d	a	-	G	a	✓
Assigned key	C	d	a	-	G	a	e
Number of cases	190	30	24	0	11	114	1
Number of exceptions	0	0	0		0	5	0

Table 4.1. Possible keys after comparing the final chord with the key signature. Here, the key signature is assumed to be the C major key signature without any accidentals.

The stress which has been put here on the need for a reliable key-finding method is grounded by the fact that in order to do key inference based on pc set chains, the normalising relation of pc set chains to a key has to be as reliable as possible. Altogether, the following tables summarise main properties of the database:

Number of pieces	386
Number of doublets	61
Number of pieces in major	201
Number of pieces in minor	185
Average length of pieces in beats	$26 \leq 64.21$ (average) ≤ 426

Average length of pieces in pcsets (using dense segmentation)	$30 \leq 95.53$ (average) ≤ 882

Table 4.2. Properties of the database

	major	minor
C	21	8
C#	0	0
D	27	27
Eb	9	0
E	10	22
F	23	2
F#	0	8
G	53	43
Ab	1	0
A	34	48
Bb	23	2
B	0	25

Table 4.3. Numbers of pieces in each key

5 Distributions of pc sets and pc set sequences

5.1 Single pc sets

The frequencies of pc sets would be expected to reflect the way pitch and harmonic relations are organised in tonality. In order to investigate this, pc set distributions for the subsets of major and minor chorales have been computed, using *dense, metrical and harmonic approximation segmentation* methods (see appendix). The distribution found with *dense segmentation* will be discussed here.

The distribution for major (table 5.1) shows various interesting properties. A rather high number of 244 different pc sets was found. As expected, the three most frequent pc sets are triads on scale degrees I, V, IV, and triads on all seven scale degrees are present within the nine most frequent pc sets which include the two dissonant pc sets V^7 and II^7 or IV^{6}_{5}. However, it is interesting that the triad on the tonic is remarkably more frequent than that on V or IV. For the V scale degree, this is due to the fact that it is split into V and V^7 which both occur within the top five ranks and would sum up to a value which is still lower, but near the frequency of the tonic pc set.[10] Furthermore, the distribution favours a differentiation between triads as groups on the scale degrees (I,V), (IV,VI,II) and (III,VII) ordered by descending frequency. In particular, the frequency of the last group is considerably lower than that of the previous group. Moreover, three dissonant pc sets, V^7, II^7 or IV^{6}_{5} and VI^7 or I^{6}_{5} are particularly prominent. Whereas all top ten chords fit the diatonic (key profile) distribution, already the 11th and 12th rank as the secondary dominant (seventh), include the non-scale pitch class F#. This is remarkable, as both relatively high frequencies of approximately 1.9% (together approximately 3.8%, which corresponds to the 7th rank) suggest a presence of the F# pitch class in the key profile. The next non-scale pc sets are found at ranks 17 (III major) and 20 (I dominant seventh), but with considerably lower percentages of 1.12% and 0.95%.

In comparison to the major profile, the minor profile (254 different pc sets) reveals aspects of a different tonal organisation. Remarkably, here the parallel major triad (III) constitutes the second most frequent pc set, which reflects a strong affinity of minor keys to their parallel major which is

10 But nota bene that incomplete versions of these chords are not included in this count due to their ambiguity as discussed above.

further underpinned by the presence of their respective dominants at the 3rd and 4th rank (V,VII). Similarly to the major profile, the fourth scale degree is frequent, while the sixth is less so though still frequent. However, a grouping of the triads appears to be not straightforward. The triad on the second scale degree is remarkably rare.

A double-logarithmic plot (diagram 5.1) of the relative pc set frequencies in major and minor shows that the top frequencies are differently distributed. Particularly, the top two ranks are particularly less frequent in minor than in major, whereas lower top ranks tend to be slightly more frequent in minor than in major. This, as well as the fact that triads involving non-natural scale pitches are particularly frequent in minor, underlines the theoretically well-known flexibility of the minor mode and its greater variety of related triads. Major and minor versions of the scale degrees V, IV, I are found to be prominent, as well as the major and diminished version of the triad on the seventh scale degree. Altogether, this confirms the a rather flexible minor scale/flexible key profile, particularly concerning the sixth and seventh scale degree pitch class, which may be equally flattened and natural. In the case of scale degree VI, this might stem from the still frequent occurrence of the Dorian sixth, or from frequent secondary dominants to the parallel major (III). In part, these results revise Aldwell & Schachter's (1989) characterisation of frequencies of triads in minor (table 5.2), as particularly, the diminished triad on scale degree VII is relatively infrequent, whereas the major triad on IV is rather frequent.

Pareto diagrams (diagram 5.2) show that the 10/13 most frequent pc sets (comprising triads on all scale degrees) in major/minor constitute 58.9%/59.88% of all pc sets. This appears to be an interestingly small number and highlights a strong and cumulative presence of dissonant pc sets, which might be understood, in one possible interpretation, as a peculiarity of Bach's densely ornamented chorale style. A statistical comparison to another contemporary chorale style, say that of Telemann, would be indicated here, but will be reserved for future research due to the limitations of space. In detail, table 5.3 lists pc set genera and their frequencies for major/minor. Altogether frequencies of consonant pc sets and dominant sevenths sum up to 12545/11731 in major/minor, whereas the sum for the remaining pc sets is 6222/6469, which forms a proportion of 0.6668:0.3332/0.6446:0.3554. This proportion, which is close to 1:2 for both modes, and the great variety of different (dissonant) pc sets underpins the density of a large number of rather individual passing phenomena. Moreover, it can also be seen that the modes appear to be rather flexible in terms of modulations as 9/10 different major, 8/11 minor triads and 7/9 dominant seventh chords appear in major/minor.

These results about pc set frequencies link with other quantitative and psychological approaches towards tonality. Budge (1943) carried out an early extensive study of chord frequencies within a large sample of a great variety of musical styles, which contained nearly 66000 chords. She focused on chords of the 7 scale degrees and counted different forms of chords, including added sevenths and ninths as instances of the same scale degrees. As might be expected, her results are confirmed by the pc set distributions computed here, and correlate highly significantly with 0.99/0.86 (see details in table 5.4). Interestingly, the frequencies of triads on each scale degree correlate well (0.94) with the key profiles for major and minor which have been proposed by Krumhansl (1990) on the basis of data from probe-tone experiments. In particular, the peculiarly strong presence of the relative major in the minor keys (significantly stronger than the relative minor in major keys – which represents the tendency towards the dominant key in major) is represented in both distributions. These correlations may be interpreted in terms of surprisingly strong fundamental interdependencies of tone and chord distributions. From one perspective this is perhaps less surprising, as chords, of course, contain tones and the distribution of the former can be expected to depend in part on that of the latter; but the apparent dependency of tone distribution on that of chords cannot be explained in the same way.

5.2 Analysing sequences of pc sets

5.2.1 Pc set transitions

The next logical step from studying single pc set distributions is to study simple progressions of pc sets. Using the *dense* and the *harmonic approximation segmentation*, 2320/2509 and 1022/1063 different progressions for major/minor have been found. Compared to the large number of different pc sets these are rather small numbers reflecting fact that the majority of rare pc sets are contained within certain individual contexts (of course all pc sets have at least single occurrences). Furthermore, the majority of pc set progressions involve a smaller number of 'highly active' pc sets, which can be seen in the following coloured scatter plots (diagram 5.3; pc sets on the Y axis progressing to pc sets on the X axis, the colour represents the logarithm of the frequency). It appears that, roughly, the progressions tend to aggregate around the 50 most frequent pc sets. Zooming into these 50-by-50 subdiagrams, it can be seen that indeed the 10 most frequent pc sets (the scale degree triadic chords) dominate most of the 'traffic'.

The data of set transitions found connects to other studies concerning the cognition of chord progressions. Bharucha & Krumhansl (1983) performed an experiment in which all possible pairwise progressions of diatonic chords in C and F# major were judged for their fit (see table 5.5). Piston (1941/1978) provides a table of usual root progressions (table 5.6) which is used as a central reference for various studies on chord progressions. However, his judgments are largely intuitively founded and are not based on any replicable quantitative method. Krumhansl (1990) correlates the results from her 1983 study with a quantification of Piston's table (using the values 1,2,3 for frequent, less frequent and rarer progressions) because “no suitable tabulation was found in the literature containing quantitative values” (Krumhansl 1990: 194). She found a highly significant correlation of 0.53 (with $p<0.01$ and 40 degrees of freedom) between her and Piston's results, which appears to suggest connections between cognitive and music-theoretic properties of chord progressions. This study should be relatable to the present research. The computed progressions between diatonic chords have been compared both with Piston's and Bharucha & Krumhansl's results. The table of progressions was compiled from the harmonic approximation segmentation. The data displays a strong and statistically significant correlation with the cognitive data.

Schmuckler (1989) performed one experiment investigating expected harmonic progressions in tonal contexts and found some similarities between his results and Piston's table. These appear to imply that Piston's table may be relatively appropriate for predicting expectations of events which "often" and "sometimes" follow, but not in terms of "seldom" or not considered chords.

However, from the perspective of describing tonality, a preliminary discussion of the relevance of a transition/2-gram analysis is particularly relevant. The distribution of single pc sets has been studied above and found to correlate with Krumhansl's (1990) key profiles. Hence, it can be asked what extra information a description of pc set pairs adds to the distribution of single pc sets – or in other words, if the distribution of pairs could be predicted by the mere distribution of single pc sets, no new insight into tonality would be gathered. Therefore, a method following Conklin & Anagnostopoulou (2001) was employed to explore the significance of 2-grams. Its core idea is to compare the frequencies of sequences found in the corpus with sequences found in a random corpus. For the purpose of this study, a random corpus of pc set sequences, containing the same number of major and minor pieces of the length of 96 (respecting the average length in the chorale corpus), was created from the distribution of single pc sets found with dense segmentation (see appendix).

As expected, one can see that 2-grams cannot be generally reproduced by the single distribution. Nevertheless, the hierarchy of pc set pairs is rather close to that in the chorale corpus and, in particular, the most prominent sequences are predicted. However, their overall frequency tends to be lower than in Bach's chorales – which suggests a weighted preference for certain progressions, although their overall distribution is similar (see below). But most importantly, whereas the pairs in the random corpus tend to be rather symmetric, this property is not shared in the chorale corpus. Important progressions tend to be directional. This confirms the common intuition of music as a goal-directed process in a quantitative way (which carries less Western music theoretical baggage) – and may fit well to an overall picture of music from a implication-realisation perspective (Narmour 1992) and directedness (Schenker 1979; Larson 2004). This asymmetry further accords with a directional asymmetry proposed by an interval-periodicity approach (Woolhouse & Cross 2004).
Generally, the transition statistics (for major) show some general similarity to Piston's judgments, but nevertheless there are some differences about single transitions, asymmetries in transitions and the underestimated VII scale degree. Moreover, the tables show strong numeric differences for the frequencies, which may suggest to quantify Piston's table by exponentially increasing numbers rather than by values of 1,2,3. This has been realised (see tables 5.7b,c) and such a quantification of Piston's table correlates significantly (0.54, $p<0.01$) with corresponding pc set frequencies in the corpus. Summarising, Piston's table significantly fails to represent the strong discrepancies between

asymmetries in progressions such as I-V/V-I and may have to be refined in several details (also the underrepresentation of the VII scale degree chord), but generally it correlates well to empirical data. Similarly, Bharucha & Krumhansl's (1983) results show a statistically highly significant correlation of 0.62 ($p<0.01$) with the empirical data.
In conclusion, the significant correlations between statistical and cognitive properties of chord progressions may be interpreted to correspond to the overall framework in the spirit of Meyer, outlined in the introduction.

5.2.2 Larger sequences of pc sets

To move from analysing pairs to analysing pc set n-grams is not straightforward, as there are too many perspectives from which larger sequences may be analysed. This study will only focus on some overall aspects of n-grams surrounding the dichotomy between standardisation and individuality of pieces; in other words, the focus will be on the extent to which n-gram statistics where $n>2$ may be informative about the structure of the corpus. As observed before, the number of different 2-grams was rather low compared to what would be predicted from all combinations of single pc sets; the proportion stands at around 1:10. In respect of the proportion of different 3-grams to 2-grams, the proportion only is around 1:3. Accordingly, it may be of interest to study how the number of different *n*-grams found changes (the proportion converging to 1) and to which extent the entire corpus can be predicted with *n*-grams.

Individuation analysis

Accordingly, an experiment on the number of different chains found within the corpus of the chorales (without separating major and minor chorales) was performed for $n = 1,2, ..., 24$ (diagram 5.4, blue curve). The red curve indicates the number of pc set sequences which were found only once and one can see that as *n* grows the number of these sequences increases rapidly. With respect to the distance between the red and the blue curve, one may observe that even for larger *n* (with values above 15) numerous repetitions of pc sequences were found. This finding led to the introduction of another variable to measure the number of sequences which occur more than once and in at least two different pieces (its curve is plotted in green).

Subsequently, the difference between the blue and the green and red curves may thus be interpreted in terms of repetitions of sequences within single pieces. The relatively slow decay the green curve, however, can only be interpreted by taking into account the mentioned existence of a number of rather similar harmonisations of the same chorales (such as “Jesu meine Freude”, “Wer nur den lieben Gott lässt walten”), which still are too different to be considered as doublets (see below). These results, however, may imply for this and other research that the patterns of longer pc sets sequences might be (to a greater or lesser degree) biased by progressions from these chorales.

The number of single sequences, which strongly increases until it approaches the empirical maximum of sequences (where nearly each sequence occurs only once), to a certain degree represents the individuality of sequences of the pieces. This is because they describe unique sequential contexts which will always remain and become larger unique contexts with increasing *n*. Hence, the increasing individuality of n-grams can be described in terms of the decomposition of sequences occurring several times into sequences occurring less frequently until they only occur once. These relationships can be studied in terms of the proportion of the overall number of sequences found against the number of unique sequences (diagram 5.5). The number of sequences occurring more than once yield information concerning the sequences in terms of tonality because a high proportion of less frequent sequences describes the data sparsity of the database rather than either actual frequencies or 'real' probabilities of sequences. A set of an enormously high proportion of individual sequences may predict nearly the entire corpus, rather than describing new combinations. For an extreme instance, two *n*-grams for a piece or a sequence of length *n*+1 may just describe the whole sequence and not bear 'new' information. Consequently, an informative description of sequences and patterns may want to avoid these cases of a too high proportion of unique sequences. Accordingly, an experiment was performed in which predictions based on *n*-grams were compared with the actual continuations for each piece. Table 5.8 shows how many percent of the corpus have been predicted correctly. Altogether, this and diagram 5.5 may give information on reasonable values of *n.* After all, values of *n* at most no greater than 6 appear to be useful here.

n	Mean percentage	Standard deviation
2	19.82%	6.41%
3	33.03%	6.81%
4	51.21%	7.82%
5	71.38%	8.84%
6	86.56%	6.20%
7	94.26%	3.95%

Table 5.8. Mean percentages of prediction of the database using n-grams (dense segmentation)

From another context of music information retrieval, Yip & Kao (1999) developed a similar diagram for melodic sequences from a different corpus containing a larger stylistic variety. As it can be seen (diagram 5.6), the behaviour of counts of melodic and pc set n-grams is extremely similar – which might favour an explanation of this as a more general property of n-grams in larger databases – which are governed by a Zipfian distribution of their elements (see below).

Distributions of pc set sequences

In computational linguistics, it is a well-known property of larger corpora that the frequencies of words are related to their rank, which is known as Zipf's law. In detail, the number of words which occur exactly *i* times is roughly described by $c(i) \sim 1/i^k$ (and more precisely, by $f(i) = (a/i^k)b^i$ (Simon 1955)) for appropriate constants *k* (which often approximates 2), *a* and *b*. One pioneer study by Zipf (1935,1949) proposed this property in an exhaustive word frequency count in Dickens's "David Copperfield" (see diagram 5.7). Accordingly, Zanette (2005) performed a similar study on pitch distributions in four different works of different styles by Bach, Mozart, Debussy and Schönberg and found a striking fit to a refined version of Zipf's law, following Simon (1955). Similar note distribution properties are found in Bach's chorales though the data is sparse. (diagrams 5.8 & 5.9). The present work will complement and extend these approaches by studying sequence frequencies within the chorale corpus.[11] (The distributions within single pieces are relatively sparse, but will be discussed shortly below). Diagram 5.11 shows rank-frequency plots on a double-logarithmic scale for pc set *n*-grams for n=1,2,3,4,5,6. It appears to display a good overall similarity to Zanette's and Zipf's curves (Diagram 5.10).

11 From another context, the discourse about context of the application of 1/f-noise to music (compare *Voss & Clarke* (1975; 1978)) is also relevant to this.

However, one can observe that the first ranks of the *n*-gram curves decay at a rather slower rate, which may speak for a larger number of relatively frequent elements necessary to disambiguate context. For example, a tonal region is not definable by means of a single chord but requires at least three events to provide unambiguous key information. However, this finding is not yet clear in its implications; it be compared against results found within the random corpus. (diagram 5.12). These show that even randomly generated 'pieces' from the initial Zipfian distribution display roughly the same behaviour for larger sequences – thus, distributions of larger pc set sequences may to a larger extent already be caused by the properties of the Zipfian distribution of single pc sets which is organised into random or non-random pieces.

Nevertheless, an overall interpretation of these results has to be made carefully. For the text-based case, Li (1992) claims to have found that randomly generated texts show a Zipfian distribution. Although this result must be interpreted cautiously, too, as its significance strongly depends on the method of generation, it nevertheless begs the question of whether the distribution is a 'natural', rather tautological, statistical outcome found in larger databases or if it is indeed based on basic properties of human-made codes and communication. Zanette (2005) interprets the general as well as the special (curved form of the graphs) by an idea following Simon (1955) who proposed a contextual model of explanation. According to this model, throughout the temporal unfolding of a text/piece, elements (words/tones/chords) which already have occurred are more likely to occur again than new events. Simon's mathematical model of this relationship fits Zanette's graphs nicely. Similarly, it does not appear implausible that the establishment of a tonal frame of reference favours certain chords so that they are much more frequent than others (which can be seen on single-piece distributions below). These curves may add up, ultimately resulting in the overall distribution. Yet this might not provide sufficient ground to favour a deeper relationship between music and language which associates tones/chords and words. Tones, as well as chords, fulfil different functions in music from the functions fulfilled by words in language. Particularly, the most common tones/chords provide some fundamentally supporting frame of reference for tonality (according to the pitch or pc set key profiles) whereas most frequent words in texts (such as “the”, ”of”, “and”) barely communicate any content-context. For lower ranks, once words of content-relevance are reached, the similarity might be better, though still possibly different in nature. Moreover, most strikingly, meaning of music is not exhausted in tonal terms, but in bodily, cultural terms (Cross 2005 forthcoming).

The distribution within single pieces

Some properties of the overall distribution of pc sets occurring throughout the corpus were discussed. A short complementary study of pc sets within single pieces is detailed. Generally, the pc set distribution within a piece is governed by very few governing pc sets representing the 'anchoring' tonal scale degrees. Much less frequently represented are a larger number of rare pc sets which tend to be rather individual for a given piece. This may represent Zipfian behaviour as well (see diagram 5.8), which cannot be claimed due to data sparsity. The distribution from one sample is given in Table 5.9. Throughout the whole database, the average number of different pc sets in the pc profiles is 30.2.

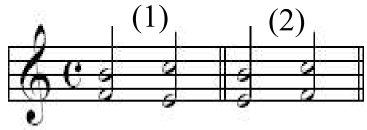
These pc set distributions may play a role in identification of pieces, which may well prove applicable in music information retrieval contexts. It appears that the pc set profile of each piece is an identifier of the individual piece, which is not necessarily the case. In a small-scale practical application, a pairwise comparison of pc set profiles was run for the whole corpus. No cases of entirely different pieces with similar pc set distributions were found, but all doublets and virtual doublets were identified and confirmed. Furthermore, very similar, but distinct versions of a chorale yielded a pc set distribution similarity of around 80%. Hence, this may outline an alternative tactic for finding doublets.

Common sequences

Comprehensive lists of *n*-grams have been compiled for n=3, 4 using *dense, metrical* and *harmonic approximation* segmentation methods. In the case of dense segmentation, the most common *n*-grams (see appendix) revealed a striking overall dominance of V-I cadential contexts above other contexts. This fits with the 32 shortest significant patterns found in Conklin (2002) who also uses a Bach chorale corpus. Consequently, this raises an unambiguous picture of tonality as essentially grounded on the cadential V-I relationship which fundamentally clarifies a key. In the case of the harmonic approximation segmentation method, most cadential patterns are reduced to simpler identical forms, which in turn reveals other frequent non-cadential patterns. Accordingly, most patterns are strongly directed and therefore differ significantly from patterns from the random corpus (compare appendix). However, a detailed analysis of patterns found in larger sequences would give enough material for an extra study, which may link to results by Eberlein (1994), but exceeds the limits of space for this project. In this context, potential future research may perform a study with earlier music to investigate factors influencing the evolutionary emergence of Western tonality which might help elucidate Dahlhaus's (1967) study on the origins of Western tonality in quantitative terms.

6 Key profiles and key induction

One essential and indispensable basic structure of music and its cognition is key. Before harmonies and successions of harmonies can be related to their function within a harmonic context, they have to be structurally referred to a key underlying the context. Hence, a cognitively relevant approach to musical structure has to account for this. It is unclear how complex the cognitive mechanisms employed are, however, at least some simple factors appear to play an important role. Krumhansl (1990) argued for the relevance of pitch profiles for key induction. These probe-tone experiment based key-profiles rank pitch classes by degrees of importance for a key. This, in connection with the results from the previous chapter, suggests that empirical pitch class frequencies within a piece relate to their relevance for perceived tonality. Correlating a small number of tones from a piece whilst processing was suggested to yield an acceptable approximation of the underlying key. Temperley (2001) suggests improvements to some shortcomings of Krumhansl's profiles which enhance the algorithm's performance. Moreover, Temperley (2004; 2002) proposes statistical probability-based Bayesian approaches to key finding, where sets of notes presented are related to possible keys. However, both the key-profile method and Temperley's Bayesian method embody significant shortcomings: straightforwardly tone-based key-statistics disregard important pieces of information such as relationships between notes or other cues such as metrical position/weight of single notes, which would probably yield improvement and very likely afford a degree of cognitive relevance. But more importantly, vertical co-occurrences are not taken into account. Temperley is aware of this problem when he gives the following example (6.1): If the notes are played in the first order a clear estimation near C-major is possible, but once they are played in the second order it is rather unclear, and the tonal implication of the last case barely exists. However, an entirely tone-based approach towards key finding will treat both cases as equal and will be too rough.



example 6.1

Within that frame of reference, the distributions of pc set sequences discussed in chapter 5 can be interpreted from another perspective, as well. They do not only describe features of tonality in terms of local organisation of vertical pitch class co-occurrences, the patterns and their frequency also directly characterise one distinct key, namely the key which the pieces have been normalised to. In this case, the *n*-gram sequences make *n*-gram profiles that are characteristic of C-major/minor. This property, then, may be applied for tasks involving the determination of the most likely or most appropriate key of a given context. A certain pc set sequence, as well as a concrete distribution of pc sets can be correlated to all 24 (transposed) versions of the key profile in order to calculate the best matching key(s).[12]
One particular property of this method is that the weights assigned to certain pc set sequences strongly influence the result, which is an interesting advantage over a method which focuses on scale-implications. For instance, from such a perspective, sequences like G-a-C, or C-G-C-G would be rather equally assigned to C-major or G-major. However, applying the pc set profiles which reflect information distributed across all Bach's chorales, a clear tendency towards C can be found in both cases.
Even single chords do favour a certain key. That a single major or minor triad results in an interpretation as tonic within that key is unsurprising, but that C-D-Eb-G, D-F-Ab-C or Eb-G-B-D already distinctively distinguish c-minor may be a useful result. In particular, these 'associations' which may not even involve (distinct) tonic features may link to perspectives from connectionist approaches to music cognition (such as Bharucha & Todd 1991, Leman 1992).
An important difference between the pc set based and the pitch class key profiles, is that the former incorporates more information on relationships of pitches in tonality, which involves a much larger number of elements in its characteristic distribution. In terms of precision, this may be an advantage as the larger distribution contains more information to be employed than the simple pitch class profile, but it is gained by the cost of generality. For, though the extra information may render the *n*-gram profile approach successful in many cases, it will expose restricted applicability in cases where the vertical structure is changed, such as extended harmony or jazz chords. The key profiles need to be compiled from a corpus from the particular style.
Furthermore, the length of the sequences has to be chosen carefully in order to keep significance, for the information on key for each *n*-gram converges rapidly from generality towards single pieces. However, it is important to keep in mind that the key-profiles are built from the global key of the entire pieces. Thus, the profiles also incorporate information in modulation structures – and thus rather represent an overall distribution.

12 Practically, not the key profiles, but the pc set sequence in question will be transposed.

Moreover, the advantage of extra information also brings problems for the method. There is a large number of pc sets and huge numbers of *n*-grams. Therefore, to be generally applicable, the method has to face the problem of data sparsity as it may happen that input includes pc sets/sequences which are unknown to the profile. As long as the query involves the same sequences that have been processed before, this is not a problem. But due to the Zipfian distribution of pc sets in pieces discussed above, unknown pieces will relatively certainly contain *n*-grams which have not been processed before. Thus, for general cases, the method has to deal with data sparsity (various methods are described in Pearce & Wiggins (2004)). It could for instance successively cut *n*-grams into *n*-1-grams until they are found or can be combined with classical pitch class profiles. Alternatively, input sequences could be segmented/reduced into sequences which can be found. However, the sparsity problems would not be solved by just enlarging the database as due to the Zipfian distribution of the elements, the size of the database would have to grow exponentially in order to reasonably increase the key profile data set. The implementation and fine-tuning of these possibilities, as well as a comparison with other key finding approaches, is potential for future work. Generally, the potential of the method stands and falls with the quality of the initial normalisation of the database, and with the amount of information included in the key profile. For instance, including further information like duration or metrical weight into the profile may be further potential for improvement.

Altogether, this method of associating a key to a given *n*-gram links to a classical case of a classification problem, in which a machine learning algorithm may be trained by a corpus. Hence, the application of more advanced machine learning methods will be explored in future research.

Another interesting perspective, for both the sequence analysis as well as the key finding, would be to respect differences of keys in that method. Instead of using the same profile for all 12 possible keys, different key profiles could be computed for all keys independently. Although, this may require a distinctively larger database, interesting results might be yielded describing individual differences between keys – which would have promising applications for music from before the establishment of modern tonality, which treats each key equally.

7 Dynamics of attention: ambiguity, revision and expectation

One essential feature of musical attention is its temporal dynamics. Frequently, music analyses implicitly assume static interpretations of musical elements. But throughout the actual process of listening, structures such as metre, grouping, harmonic functions or key implications are not necessarily statically assigned to past structures of the piece; they may change dynamically/retrospectively while the piece unfolds. Generally, the dynamics of diachronic music processing involves two aspects: interpretations of heard material and expectations, both of which may change across time.

Mainly in respect of melody, Eerola (2003), following Bharucha (1987), distinguishes between three kinds of expectation: data-driven, schematic and veridical expectations. The first accounts for cases where expectancies rise from general perceptual principles, like gestalt principles of good continuation. The second concept, in contrast refers to expectations which rise due to particular, culture-dependent knowledge of typical common stylistic patterns. The case once a piece is already known is entirely described by veridical expectation.
In terms of dynamics of interpretation of heard material, two particular phenomena are interesting: revision and ambiguity.

Ambiguity characterises a case where interpretation is not distinctively possible either because too little information is given or the information favours several possibilities equally. The chord sequence C-G-C-G from above might appear as ambiguous between C-major and G-major. But similarly, one has to be careful as to the assumptions (on distributions) on which such cases reside. For instance, Agawu (1994) describes the nicely imaginary example that 'strictly' speaking the beginning of Beethoven's fifth symphony would be highly ambiguous, as the two presented pitch classes G, Eb would allow 3 major and 11 minor interpretations. This is true – from a perspective which involves a 'flat' key-profile (like the Longuet-Higgins/Steedman (1971) key finding algorithm, too). However, a weighted key profile like Krumhansl's reduces the number of potential keys drastically. Similarly, the chord sequence above scores higher for C-major than for G-major employing the pc set distributions above.

Revision is a phenomenon which involves the reinterpretation of an already assigned, preferred analysed structure due to contradicting evidence. In often-cited linguistic examples like “the horse raised past the barn fell” or “the old man the boats” initially most plausible interpretations of “raised” as verb or “man” as noun have to be reinterpreted by less intuitive interpretations as past participle or verb once the diachronic interpretation of the continued sentence becomes untenable. Jackendoff (1992) demonstrated the reality and relevance of revision for the musical case by an example of different metrical interpretations depending on different potential continuations of the beginning of a chorale by Bach. Temperley (2001) gives cases for revision concerning the parameters metre, harmony and grouping. Revision is not a rare event in harmonic structures. A common example of harmonic revision is simply given by modulations. Indeed, modulations frequently involve pivot chords which fit both the original and the modulated key and are subject to (also functional) reinterpretation. Moreover, Aldwell & Schachter (1989: ch.32) discuss revision in the context of enharmonic modulation.

Example 7.1

However, mere modulations appear not to show the striking garden-path effect of the sentences before. Accordingly, the beginning of Schumann's song “Am leuchtenden Sommermorgen” (in “Dichterliebe”, op.48 No.12, harmonic reduction in example 7.1) appears frequently as a key example for revision, because the initial (broken) chord will be inevitably heard as dominant seventh which is effectively revised as German sixth with bar 2. But nevertheless, this example is not categorically different from the case of a common modulation which reinterprets the functional implications of a pivot chord. The difference may rather consist in the surprise, the amount of necessary reinterpretations and the harmonic distance which may be connected to the rareness of the case. Similarly, the difference between modulation and local tonicisation may be just gradual, depending on the degree of stability of the introduced key.

Schumann's song also illustrates another phenomenon. At the beginning of the next phrase (bar 6-7), the same chord progression from the beginning is repeated. Also in bar 8-9, another (enharmonically respelled) instance of the initial chord is repeated. But now, the garden-path effect consists in the common V^7-I continuation which traps the expectancy of a repetition of the more rare initial progression. However, the way this expectation works seems not to fit into the trichotomy given above. It demonstrates an instance where an on-line set of expectations is constructed, revealed and abrogated. Thus the example is not strictly schematic (stylistic) knowledge, it is rather a dynamic schematic short-term expectancy – which gives a musical piece of evidence for the reality and relevance of such on-time expectancies.

In terms of expectation, it is not clear how the above concepts relate to harmony. Whereas veridical and style-based schematic knowledge are unproblematic, data-driven expectation is not. For it is not clear how data-driven harmonic patterns would appear without being stylistic patterns. Gestalt based continuations are relatively straight-forward in terms of melody, but in terms of harmony it is not clear, for instance, if the bass would had such a prominent role in harmony cognition that a sequence of harmonies based on chromatically or fifth-wise descending bass could be accounted for by gestalt principles, although multiple other pitches are involved.

7.1 Sliding window key induction

A small and simple model of the discussed theoretical concepts can be built for the range of key induction. A piece can be processed using a sliding window approach and for each of these windows, its key (approximation) and/or a prediction about the following pc set/key can be computed. In this respect, the key induction method from the previous chapter can be applied easily. Using different sequence lengths *n* for the key estimation, the relevance of contexts of different size can be studied. But even so one has to be aware that with larger values of *n* the data becomes sparse and insignificant fairly quickly. Concerning the particular dimensions of this project, sequence lengths of the value 1 (single pc-sets) to 4 turned out to be practicable.

From a cognitive perspective, a sliding window technique appears slightly more adequate than Temperley's (2001) similar approach, which computes an entire key analysis from the very beginning to the current time point. Human attention can be assumed to be restricted by temporal limits in relation to short term memory and processing capacity. This and the fact that in music psychology there is not much known about the relevance and nature of large scale features, should motivate a sliding window approach as a viable partial model.

This approach has been applied both with *dense segmentation* and *harmonic approximation*. Example 7.2 demonstrates the way this model functions and displays the sliding windows for all lengths. The key assigned to each sliding window is marked by the symbol assigned to the last pc set. Hence, a symbol in the third row indicates that a context of the previous 2 chords is taken for the key association. Example 7.3 illustrates a practical application of this model and may yield some implications for the role of context in key induction. First of all, once the sliding window consists only of one pc set, as expected the implied key shifts more or less with each pc set. Once more context pc sets are taken into account, the estimated key stays more and more stable, and clearly, with larger *n* it must converge towards the key of the piece, due to the way the key is computed. So, once larger contextual windows are employed, the associated key tends to expose a certain 'inertia'; changes which occur on first and second level are carried through much more slowly. Interestingly, contextual changes tend to be slightly delayed in their effect. This can be seen in example 7.4 from "Jesu, Leiden, Pein und Tod": the sudden modulation (to $b_\flat$-minor) affects the 4^{th} level last of all. The effect may be explained by the fact that some time steps need to pass until the influential context contains fewer pc sets from the previous key.
Moreover, it turned out that already in the case of 4-grams sequences occur which just occur once. In order to avoid problems caused by data sparsity, a threshold has been introduced to rule out cases which occur only once and thus are not representative of the dataset. Additionally, another threshold of minimal significant difference has been introduced to mark cases as ambiguities where two or more keys are assigned with fairly equal scores (more than 90%). In the diagrams, ambiguous cases are marked by "?", whereas sparse data is marked by "*". For further illustration, simple pc set predictions have been computed for some examples, building on the *maximum likelihood* estimation described above. Key predictions have been produced by evaluating the context including the predicted pc set with the key induction method.

The sliding window approach can be compared to classical cases of music analysis. Daniel (2004) remarks the double case of an interrupted cadence at the example 7.5, which finds an adequate interpretation in the underlying text "falschen Tücken" [false deceptions]. First, a cadential context towards Bb-major is established and then abrogated by D^7 which itself sets up a revised cadential context towards g-minor which is abrogated a second time by Eb. Then, a continuation to C^7 alludes to F-major before the piece finally reaches the Bb-major cadence. A comparison of this context with the results of the model shows that on a level of bi-grams this interpretation finds its confirmation whereas already at the levels 3 and 4 the contexts turn out to be weaker.

An example which would be a parallel case to the Schumann song discussed is the chorale “Es ist genug” (B91; example 7.6), which exposes a number of rare and exceptional phenomena. The first phrase already presents two cases of revision. The established A-major context is reinterpreted to B with the third chord F#, which also involves an exceptional accented strong dissonance (which is indeed the only occurrence of this kind throughout the dataset). The next chord $G\#^7$, however, effectively destroys the clear cadential context and forces another key revision towards c#-minor. After the following two phrases which cadence on A and E, another parallel instance of the initial progression is exposed, which however, does not repeat the first revision but then repeats the deceptive cadence character with a diminished D# triad. An analysis with the model confirms these effects. Furthermore, concerning the above concept of short-term on-line built expectancy, it will be a future perspective to include an instance of an on-line maintained individual distribution for a piece into the model, in a form such as Conklin (1990) proposes. However, it is yet unclear how such an extension should be weighted against the stylistic distribution from the whole database, because the sparse data may distort some results in normal cases. Possibly, rare exceptional (low-probability and high standard deviation) cases could be treated with an on-line distribution.

Another telling example is given by the beginning of “Jesu, Leiden, Pein und Tod” (B194; example 7.7). The computed “key fields” display an interesting key ambiguity between Eb major and c minor. It can be seen that, in terms of statistics throughout the body of Bach's chorales, already the transition Eb – Bb – Cm suffices to produce a likely interpretation of c-minor. This underlines the tight and ambiguous relationships minor and relative major keys may have. But it also may lead to implications about modulation. Though the model had turned out not to modulate too easily on higher levels, this behaviour suggests an instance of an interesting modulation pattern, as there is no leading tone/dominant involved. In this case, for instance the pc set C-D-Eb-G appears to suggest c-minor so distinctively that it overrides the Eb-major context.

This result has implications in respect of music theory. The classical notion of modulation, e.g. following Schönberg (1969; 1978), necessarily includes the presence of a dominant or leading note context, which will have a stabilising role. But here, a case of music practice is found where change of key is initiated and stabilised by other means without central participation by a dominant.

Similarly, it might challenge the “primacy hypothesis” in Brown et al. (1994), which postulates key cognition to be governed by rather statically by the very first harmonic evidence presented which only changes once sufficient counter-evidence has emerged. A probabilistic interpretation of key, may be a more flexible alternative, and may encompass ambiguities and under-determination, as well. It will have to be a matter of psychological evaluation as to whether this holds for key cognition, but it raises an interesting potential conceptual development.

7.2 Discussion

The model presented here computes key estimations from a frequency/probability basis for a smaller sliding window of context. Generally, the resulting key estimations show that basic chord/pc-set statistics and relatively small context are already sufficient to reproduce practicable results and to reproduce some effects of attentional dynamics. However, pc set statistics turned out not to have good predictive power unless an entire individual sequence were to be reproduced (which would correspond to the case of veridical knowledge). Altogether, it is not clear in how far these aspects have implications for music cognition. The statistical data sufficient for key estimation might be relevant for schematic (stylistic) knowledge and could fit with connectionist approaches as methods such as associative networks appear to show similar behaviour and to provide well-fitting capacities of storing statistical information (MacKay 2003) Furthermore, the vagueness, but error-correcting properties of network models appear to fit the present results on prediction as well.

However, there are differences to the model Jackendoff (1992) proposed. Whereas Jackendoff suggested a number of concurrent-parallel processed sequential analyses, which are ruled out as soon as enough contradicting evidence is gathered, the sliding window approach would favour only one interpretation at each time[13] which may change dynamically; further, instead of active processing of interpretations, mere statistical 'association' from presented evidence yields acceptable key estimations. Moreover, within this model the constancy of a more global key is not problematic and evolves by itself from context, as soon as enough determinate information is passed over, hand by hand. Therefore, the sliding window would offer an easier interpretation than a rule-based approach, as no calculations are required and no current analyses are maintained, but only current context is processed. This does not refute preference-rule-based approaches like Lerdahl & Jackendoff (1983), although their cognitive relevance would appear to be grounded in particular functionalist models of the mind (like Jackendoff (1987) or Fodor 1983), whereas the statistical

13 It could imply more than one interpretation, if keys with fairly equal scores are not treated as ambiguous but as two interpretations.

approach (possibly like Temperley's (2004) Bayesian approach) would fit better with connectionist understandings.[14]

However, an entirely global notion of key for a piece is not part of the model's architecture – and, clearly, cannot be reproduced with it. Nevertheless, it is not clear to which extent this fits with or diverges from music cognition. Whereas classical music theory essentially builds on the notion of a stable overall diatonic framework of reference (see harmonic/Riemannian or Schenkerian analysis, or theories by Lerdahl&Jackendoff (1983) or Narmour (1989; 1992)), this conviction appears to be grounded mainly by (score-based) music analysis. But from a psychological perspective, there is not much known about this. Cook (1987) proposed a case against the relevance of a global tonal framework (though Gjerdingen (1999) strikingly criticised the methodological shortcomings of this study). However, even, if tonal or harmonic factors turned out to be local in nature, global (tonal) context could still be relevant, as being built by an interplay of other factors, such as timbre recognition, which Ian Cross suggested (personal communication). Before conclusions concerning the relationship between local and global key can be drawn, more research will be necessary.

Nevertheless, the model may be a practicable step towards understanding the manifold dynamics of musical attention discussed above. There are various ways in which it could be improved, refined or extended. Firstly, the key association fundamentally depends on the key estimation method discussed in chapter 5, which may be improved by methods dealing with data sparsity as mentioned above. Secondly, the aspects of chord prediction may be improved using better chord representations and including aspects of dynamically built on-line short-term expectations. Moreover, quantitative models of harmonic distance such as Lerdahl (2001) or Krumhansl (1990) could be implemented to measure the 'strength' of the revision.

Last but not least, it would be interesting to relate the methods and its results to other approaches on attentional dynamics. Widmer (2005) computes and analyses performance trajectories. Most related would be computational comparisons of the differences between the key inductions yielded by this model and other dynamic models such as Temperley's (2004) or Krumhansl & Toivainen's (2004) network.

14 Furthermore, it might turn out that both approaches nevertheless describe similar properties from different perspectives; besides, preference-rules may nevertheless be a rather practical way to describe a musical analysis.

8 Conclusions

In summary, distributions of pc sets and pc set sequences in Bach's chorales have been analysed employing different methods of segmentation. The profiles found essentially corresponded to characterisations of tonality from music analytical and psychological perspectives; in particular pc set distributions of the major and minor modes and the fundamental directedness of tonality have been confirmed. Subsequently, the aggregated distributions found have been directly employed in a method of key induction, which revealed that sequences may clearly indicate one key even if no distinct classical distinguishing features such as leading tones or dominants are involved. This key induction methods has been employed in a simple model of the dynamics of diachronic music attention and appeared to be able to reflect properties of modulation, revision and ambiguity.

There is much space for future directions and improvements. As discussed, the key induction method could be extended to be fully applicable for unknown pieces. Furthermore, in order to gain deeper insight into tonality and tonal patterns, metrical or duration information could be taken into account, or extra information on phrasing or (human) harmonic analysis could be used from an annotated corpus. Moreover, results derived from a pc set analysis may be compared to results from analysing harmonies or outer-voice movement. Another perspective would be to perform statistical analyses with corpora of different historical origin in order to reveal characteristic profiles for composers or to investigate large scale historical changes.

A further perspective for future research will be to apply methods of grammar learning to the pc set sequences. Clement (1988) proposed a model which is capable of learning probabilistic regular grammars – but he inadequately assumes, following Piston's progression table, that harmony is subsumed by a simple regular grammar. In particular, a grammatical analysis may touch the topic if music structure is intrinsically recursive which is of core relevance for hierarchical analyses like Schenker or Lerdahl & Jackendoff (1983). Furthermore, Tidhar (2005) discusses musical grammars and their parsing in ways which are relevant for future developments of the present approach. Moreover, results may be compared with other models of sequence learning such as Raphael (2004) who trains a hidden Markov model for harmonic analysis. From another perspective, this may be relevant for exploring structural and cognitive relationships between music and language (Patel 2003). In particular, it will be related to studies on grammar language learning such as Rebuschat (2005 unpub.), and to topics like the relevance of grammatical rules for cognition (comp. Bod 1998;

2001).

Finally, this work has found results which may motivate related psychological experiments on aspects such as key induction of sequences, the influence of (length of) local context for key perception, the relevance of global key, and the perception of dissonance in pc sets and preferences for musical segmentations/reductions. A comparison of computational and psychological results (like Witten et al. (1994)) may contribute in both elucidating human music cognition and improving computational approaches to music analysis.

9 References

Agawu, K. (1994). Ambiguity in tonal music. a preliminary study. in: A. Pople (ed.). *Theory, analysis and meaning in music.* Cambridge. 86-107

Alamkan, C., *Birmingham, W.* & *Simoni, M.* (1999). Stylistic structures: an initial investigation of the stochastic generation of music. Technical Report CSD-TR-395-99, Electrical engineering and computer science department, University of Michigan.

Aldwell, E. & *Schachter, C.* (1989). Harmony and Voice Leading. Harcourt Brace Jovanovich, San Diego, second edition

Allan, M. M. (2002) Harmonising Chorales in the Style of Johann Sebastian Bach. Master's thesis, School of Informatics, University of Edinburgh, UK.

Allan, M. & *Williams, C. K. I.* (2005). Harmonising Chorales by Probabilistic Inference. To appear in: *Advances in Neural Information Processing Systems* 17. MIT Press

Ames, C. (1989). The Markov process as a compositional model. A survey and tutorial. *Leonardo*, 22(2):175-187.

Baily, J. (1985): Musical Structure and Human Movement. In: Music Structure and Cognition (ed. Howell, P. & al.), London: Academic Press, pp. 237-58.

Barthelemy, J., & *Bonardi, A*. (2001). Figured bass and tonality recognition. *Proceedings of the 2nd Annual International Symposium on Music Information Retrieval* (ISMIR 2001), 129–136.

Bent, I. & *Pople, A*. (2001). 'Analysis'. In: Grove Music Online ed. L. Macy (accessed 5 July 2005). http://www.grovemusic.com

Beran, J. & *Mazzola, G.* (1999). Analyzing musical structure and performance - a statistical approach. *Statistical Science*, 14 (1). 47-79.

Beran, J. & *Mazzola, G.* (2001). Musical Composition and performance- statistical decomposition and interpretation. *Student,* 4 (1): 13-42.

Beran, J. (2004). Statistics in musicology, Chapman & Hall.

Bharucha, J. J. (1987). Music cognition and perceptual facilitation: A connectionist framework. *Music Perception,* 5, 1-30.

Bharucha, J.J., & Krumhansl, C.L. (1983). The representation of harmonic structure in music: Hierarchies of stability as a function of context. *Cognition,* 13: 63-102.

Bharucha, J. J. & Todd, P.M. (1991). Modelling the Perception of Tonal Structure with Neural Nets. In *P. M. Todd, G. Loy*, (eds.). Music and Connectionism. MIT Press. 128–137.

Blacking, J. (1961). Patterns of Nsenga kalimba music. *African Music*, 2(4), 3-20.

Budge, H. (1943). A study of chord frequencies. Teacher College

Burns, L. (1995). Bach's Modal Chorales. Pendragon: Stuyvesant.

Bod, R. (1998). Beyond Grammar: An Experience-Based Theory of Language. CSLI Publications/Cambridge University Press.

Bod, R. (2001).Using Natural Language Processing Techniques for Musical Parsing. Proceedings ACH/ALLC 2001. New York, NY.

Bod, R. (2002). Memory-based models of melodic analysis: Challenging the Gestalt principles. *Journal of New Music Research,* 31: 27-37.

Clement, B. (1998). Learning harmonic progression using Markov models. Online Manuscript. http://ai.eecs.umich.edu/people/bradc/papers/545proj.ps

Brooks, F.P. Jr., Hopkins, A.L., Neumann, P.G. & Wright, W.V.(1957). An experiment in musical composition. *IRE Transactions on Electronic Computers*, EC-6(1). 175–182.

Brown, H., Butler, D., & Jones, M. (1994). Musical and temporal influences on key discovery. *Music Perception,* 11: 371–407.

Chai, W. & Vercoe, B. (2001). Folk music classification using hidden Markov models. In Proc International Conference on Artificial Intelligence.

Chomsky, N. (1957). Syntactic Structures. The Hague: Mouton.

Cohen, J. E. (1962). Information theory and music. *Behavioral Science*, 7: 137-63.

Conklin, D. (1990). Prediction and entropy of music. Master's thesis, Department of Computer Science, University of Calgary. Technical Report 1990–390–14.

Conklin, D. (2002). Representation and discovery of vertical patterns in music. in: C. Anagnostopoulou; M. Ferrand; A. Smaill (eds.). *Proceedings of the Second International Conference of Music and Artificial Intelligence.* 32–42.

Conklin, D. & Anagnostopoulou, C. (2001) Representation and discovery of multiple viewpoint patterns. In Proceedings of the International Computer Music Conference, 479-485, Havana, Cuba.

Conklin, D. & Witten, I. (1995). Multiple viewpoint systems for music prediction. *Journal of New Music Research*, 24(1): 51-73.

Cook, N. (1987). The Perception of large-scale tonal closure. *Music Perception, 5, 197-206.*

Coons, E. & D. Kraenhenbuehl (1958). Information as a measure of structure in music, *Journal of Music Theory*, 2: 127-61.

Cope, D. (1991). Computers and Musical Style. Oxford University Press.

Cross, Ian (1998). Music analysis & music perception. *Music Analysis,* 17: 3-20.

Cross, I. (2003). Music as biocultural phenomenon. In: *G. Avanzini, C. Faienza, D. Minciacchi, et al* (eds.). Annals of the New York Academy of Sciences (The Neurosciences and Music).

Vol. 999: 106-111.

Cross, I. (2004). Music as an emergent exaptation. Paper given at ESF workshop on Music & Evolution, Reading University, September 2004. cited with permission of the author.

Cross, I. (2005 i.p.). Musicality and the human capacity for culture. To be published 2005. Cited with the permission of the author.

Dahlhaus, C. (1967). Untersuchungen über die Entstehung der harmonischen Tonalität. Bärenreiter: Kassel.

Daniel, T. (2000) Der Choralsatz bei Bach und seinen Zeitgenossen. Eine historische Satzlehre. Dohr: Köln.

Eberlein, R. (1994). Die Entstehung der tonalen Klangsyntax [The emergence of tonal-harmonic syntax]. Peter Lang: Frankfurt.

Ebcioglu, K. (1992). An expert system for harmonising chorales in the style of J. S. Bach. In *M. Balaban, K. Ebcioglu, & O. Laske* (eds.). Understanding Music with AI: Perspectives on Music Cognition. 294–333. Cambridge, MA: MIT Press.

Eerola, T. (2003). The Dynamics of Musical Expectancy. Cross-Cultural and Statistical Approaches to Melodic Expectations. Doctoral dissertation. Jyväskylä Studies in Arts 9. Jyväskylä

Fodor, J. (1983). The modularity of mind. Cambridge, MA: MIT Press.

Forte, A. (1973). The Structure of Atonal Music, Yale University Press.

Gauldin, R. (1997). Harmony Practice in Tonal Music. WW Norton & Company: New York

Gjerdingen, R. (1999). An experimental music theory? In *N. Cook* & *M. Everist* (eds.). Rethinking Music. 161–170. Oxford University Press: Oxford.

Grabner, H. (1974) Handbuch der funktionellen Harmonielehre. Vol. I and II. Bosse: Regensburg

Freeman, L. C. & Merriam, A. P. (1956) "Statistical classification in anthropology: An application to ethnomusicology", American Anthropologist, 58, 464-472

Hanlon & Ledlie (2002). CPU Bach. An Automatic Chorale Harmonization System. Online manuscript, (accessed June 2005). http://www.fas.harvard.edu/~ledlie/cs/CPUBach.pdf

Huron, D. (1999) Music Research using Humdrum: A user's guide. http://dactyl.som.ohio-state.edu/Humdrum/guide.toc.html

Hild, H., Feulner, J., & Menzel, D. (1992). HARMONET: a neural net for harmonising chorales in the style of J. S. Bach. In *R. P. Lippmann, J. E. Moody, & D. S. Touretzky* (eds.), Advances in Neural Information Processing 4. 267–274. San Francisco: Morgan Kaufmann

Hiller, L., & Isaacson, L. (1959). Experimental Music. New York: McGraw–Hill.

Hiller, L. A. & Fuller, R. (1967). Structure and information in Webern's Symphonie, Opus 21. *Journal of Music Theory*, 11: 60-115.

Hoffman, T. & *Birmingham, W. P.* (2000). A Constraint Satisfaction Approach to Tonal Harmonic Analysis. Electrical Engineering and Computer Science Department. Technical Report CSE-TR-397-99. University of Michigan.

Holtzmann, S. R. (1977). A program for key determination. *Interface*, 6, 29-56.

Hörnel, D. (2004). CHORDNET: Learning and Producing Voice Leading with Neural Networks and Dynamic Programming. *Journal of New Music Research;* 33 (4): 387–397.

Huron, D. (1997). Humdrum and Kern: selective feature encoding. in: E. Selfridge-Field (ed.). *Beyond MIDI: The Handbook of Musical Codes.* MIT Press: Cambridge, MA. 375–401.

Jackendoff, R. (1987). Consciousness and the computational mind. MIT Press: Cambridge, MA.

Jackendoff, R. (1992), Musical Parsing and Musical Affect, in: Languages of the Mind. Essays on Mental Representation. Cambridge, MA.

Jin, H., & *Jagadish, H.V.* (2002). Indexing hidden markov models for music retrieval. Third International Conference on Music Information Retrieval. Paris, France. pp. 20–24.

Jørgensen, M. E. & *Madsen, S. T.* (2002). Harmonisation of Bach chorales. KBS project report. Daimi. University of Aarhus.

Knipphals, H.-J. & Möller, D. (1995). Johann Sebastian Bach, der Choralsatz: ein Lehrwerk. Möseler: Wolfenbüttel.

Kraehenbuehl, D. (1958). Information as a Measure of Structure in Music. *Journal of Music Theory*, 2: 127-161

Krumhansl, C. L. (1990) Cognitive foundations of musical pitch. Ox-ford University Press.

Krumhansl, C. L. & *Toiviainen, P.* (2001). Tonal cognition. In: R.J., Zatorre, I. Peretz (eds.). The Biological Foundations of Music. Vol. 930. New York: NewYork Academy of Sciences. 77–91

Yip, C. & *Kao*, B. (1999). A study on musical features of melody databases. Technical report, Hong-Kong university, Computer Science Department.

Larson, S. (2004). Musical Forces and Melodic Expectations: Comparing Computer Models and Experimental Results. *Music Perception*, 21 (4), 457–498.

*Leman, M. (*1992). The Ontogenesis of Tonal Semantics: Results of a Computer Study. In *P. M. Todd, G. Loy*, (eds.). Music and Connectionism. MIT Press. 100-127

Lerdahl, F. (2001). Tonal Pitch Space. Oxford Univ. Press: New York.

Lerdhal, F. & Jackendorf, R. (1983). A Generative Theory of Tonal Music. Cambridge, MA: MIT Press.

Li, W. (1992). Random texts exhibit Zipf's-law-like word frequency distribution. *IEEE Transactions on Information Theory*, 38(6): 842-1845.

Lomax, A. (1968), Folk song style and culture, Washington, D.C.: American Association for the Advancement of Science.

Longuet-Higgins, H. C. & Steedman, M. J. (1971). On interpreting Bach. *Machine Intelligence,* 6, 221-41

Luhmann, Niklas (1990). Weltkunst. in: Niklas Luhmann/Frederick D. Bunsen/Dirk Baecker (eds.). *Unbeobachtbare Welt: Über Kunst und Architektur*, Bielefeld (Haux). 7-45

Luhmann, Niklas (2000). Art as a Social System. transl. Eva M. Knodt, Stanford UP: Stanford. Orig. *Luhmann, N.* (1997). Die Kunst der Gesellschaft. Suhrkamp: Frankfurt a.M.

MacKay, D. J. C. (2003). Information theory, inference, and learning algorithms. Cambridge, UK; New York: Cambridge University Press

Maxwell, H. J. (1992). An Expert System for Harmonic Analysis of Tonal Music. In *M. Balaban, K. Ebcioglu, & O. Laske* (eds.). Understanding Music with AI: Perspectives on Music Cognition, AAAI Press. 335-353.

Meyer, L. B. (1956). Emotion and meaning in music. University of Chicago Press, London.

Meyer, L. B. (1967). Meaning in music and information theory. In Music, The Arts, and Ideas. University of Chicago Press: Chicago

Meyerson, R. (2001). Bach-Prop: Modelling Bach's Harmonization Style with a Back-Propagation Network. Online manuscript, accessed June 2005. http://www.cogs.indiana.edu/ugpapers/RobM_Q400_Bach.PDF

Narmour, E. (1989). The Analysis and Cognition of Basic Melodic Structures. The Implication-Realization Model. University of Chicago Press.

Narmour, E. (1992). The Analysis and Cognition of Melodic Complexity. The Implication-Realization Model. University of Chicago Press.

Pardo, B. & Birmingham , W. P. (2002). Algorithms for Chordal Analysis. *Computer Music Journal,* 26 (2): 27-49.

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neurosciences*, 6: 674-681.

Pearce, M. T. & *Wiggins, G. A*. (2004). Improved Methods for Statistical Modelling of Monophonic Music. *Journal of New Music Research,* 33 (4): 367-385.

Pickens, J., Bello, J.P., Monti, G., Sandler, M.B., Crawford, T., Dovey, M., & *Byrd, D.* (2003). Polyphonic score retrieval using polyphonic audio queries: A harmonic modeling approach. *Journal of New Music Research*, 32 (2), 223–236.

Pinkerton, R.C. (1956). Information theory and melody. *Scientific American*, 194(2): 77-86.

Piston, W. (1941/1978). Harmony. WW Norton & Company: New York.

Polth, Michael (2001a). Nicht System - nicht Resultat: Zur Bestimmung von harmonischer

Tonalität. *Musik & Ästhetik,* 18:12-36

Polth, Michael (2001b). Ist die Funktionstheorie eine Theorie von Funktionalität? *Musiktheorie,* 16: 319-324

Ponsford, D., Wiggins, G. & Mellish, C. (1999). Statistical learning of harmonic movement. *Journal of New Music Research,* 28: 150-77.

Raphael, C. (2002). Automatic Transcription of Piano Music. *Proceedings of the 3rd Annual International Symposium on Music Information Retrieval.*

Raphael, C. & *Stoddard, J.* (2004). Functional Analysis Using Probabilistic Models. *Computer Music Journal,* 28 (3): 45-52.

Rebuschat, P. (2005 unpublished). The implicit learning of syntactic rules: An experiment on the acquisition of a semi-artificial language. Unpublished manuscript

Reis, B. Y. (1999). Simulating Music Learning with Autonomous Listening Agents: Entropy, Ambiguity and Context. PhD thesis, Computer Laboratory. University of Cambridge, UK.

Richter, B. F. (1912). Johann Sebastian Bach. 389 Choralgesänge für vierstimmigen gemischten Chor. Breitkopf & Härtel: Wiesbaden

Riemenschneider, A. (1941). 371 harmonised chorales and 69 chorale melodies with figured bass. New York: G. Schirmer, Inc.

Rosch, E. (1978). Principles of categorization. In *E. Rosch, B. B. Lloyd* (eds.). Cognition and categorization. Erlbaum: Hillsdale, NJ.

Schenker, H. (1979). Free Composition. (tr. E. Oster. Original work published 1935). Longman: New York.

Schönberg, A. (1969). Structural Functions of Harmony, Faber and Faber: London.

Schönberg, A. (1978). *Theory of Harmony. Faber and Faber: London.*

Schmuckler, M. (1989). Expectation and music: Investigation of melodic and harmonic processes. *Music Perception*, 7, 109–50.

Selfridge-Field, E. (ed.). (1997). Beyond MIDI: The Handbook of Musical Codes. MIT Press: Cambridge, MA.

*Xi Shao, C. Xu, & M. S. Kankanhall*i, (2004). Unsupervised classification of music genre using hidden markov model. In: Proceedings of the IEEE International Conference of Multimedia Expo.

Simon, H.A. (1955). On a class of skew distribution functions. *Biometrika,* 42: 425-440.

Smoliar, S. W. (1980). A Computer Aid for Schenkerian Analysis. *Computer Music Journal*, 4(2): 41-59.

Sudnow, D. (2001). Ways of the Hand: A Rewritten Account. MIT Press. Cambridge, MA.

Takeda et al. (2002). Automatic Rhythm Transcription from Multiphonic MIDI Signals

Takeda, H., Nishimoto, T., & Sagayama, S. (2003). Automatic Rhythm Transcription from Multiphonic MIDI Signals. *In 4thInternational Conference on Music Information Retrieval.*

Temperley, D. (2001). The cognition of basic musical structures. MIT Press: Cambridge, MA.

Temperley, D. (2002). A Bayesian approach to key-finding. In C. Anagnostopoulou, M. Ferrand, & A. Smaill (eds.). *Music and Artificial Intelligence.* 195-206. Berlin: Springer.

Temperley, D. (2004). Bayesian models of musical structure and cognition, in: Musicae Scientiae (8): 175-205

Temperley, D. & Sleator, D. (1999). Modeling meter and harmony: A preference-rule approach. *Computer Music Journal*, 23(1): 10-27.

Tidhar, D. (2005). A Hierarchical and Deterministic Approach to Music Grammars and its Application to Unmeasured Preludes. PhD-thesis. Technische Universität Berlin.

Voss, R.F. & Clarke, J. (1975), 1/f noise in music and speech. *Nature,* 258: 317-318.

Voss, R.F. & Clarke, J. (1978), "1/f noise" in music: Music from 1/f noise. *Journal of the Acoustical Society of America,* 63, (1). 258-263

Widmer, G. (2005). Studying a Creative Act with Computers. Music Performance Studies with Automated Discovery Methods. *Musicae Scientiae* 9(1), 11-30

Winograd, T. (1968). Linguistics and the Computer Analysis of Tonal Harmony. *Journal of Music Theory,* 12: 2-49

Witten, I. H., Manzara, LC & Conklin, D. (1994). Comparing human and computational models of music prediction. *Computer Music Journal* 18(1): 70-80.

Wittgenstein, Ludwig (1953). Philosophical Investigations. trans. by G.E.M. Anscombe. Basil Blackwell: Oxford

Woolhouse, M. & Cross, I. (2004). The Role of Interval Periodicity in Hierarchical Pitch Perception. Given at the 8th International Conference on Music Perception & Cognition. Cited with the permission of the authors.

Youngblood, J. E. (1958). Style as information. *Journal of Music Theory*, 2: 24-35.

Zanette, D. H. (2005 i.p.). Zipf's law and the creation of musical context. To appear in *Musicae Scientiae.* Online preprint at *arXiv:cs.CL/0406015v1*

Zipf, G. K. (1935). The Psycho-Biology of Language. Houghton Mifflin: Boston.

Zipf, G. K. (1949). Human Behaviour and the Principle of Least Effort. Addison-Wesley: Cambridge, MA.

cited Internet links

http://www.jsbchorales.net/sets.html

http://www.jyu.fi/musica/miditoolbox/

http://dactyl.som.ohio-state.edu/Music824/databases.index.html

10 Appendix A

The following tables list the most 50 frequent *n*-grams for n = 1,2,3,4 which were found in the chorale corpus using dense segmentation, metrical segmentation and harmonic approximation. Furthermore, *n*-grams were compiled from one random corpus of the same size as the chorale corpus using the same segmentation methods.

For each n-gram, the tables give information on its absolute frequency (*f*), its relative frequency/probability (*p*), its frequency within the random corpus (*rf*) and its relative frequency/probability (*rp*) in the random corpus.

50 most frequent pc sets in major using dense segmentation

		f	*p*
1	(C.E.G)	3446	0.18451
2	(D.G.B)	2254	0.12069
3	(C.F.A)	1146	0.061362
4	(C.E.A)	1025	0.054883
5	(D.F.G.B)	882	0.047226
6	(D.F.A)	698	0.037374
7	(E.G.B)	433	0.023185
8	(C.D.F.A)	392	0.02099
9	(D.F.B)	374	0.020026
10	(C.E.G.A)	358	0.019169
11	(C.D.F#.A)	357	0.019115
12	(D.F#.A)	353	0.018901
13	(C.D.G)	334	0.017884
14	(C.E.F.A)	313	0.016759
15	(C.E.G.B)	295	0.015796
16	(C.D.E.G)	267	0.014296
17	(E.G#.B)	210	0.011244
18	(D.E.G.B)	204	0.010923
19	(C.D.F.G)	192	0.010281
20	(C.E.G.A#)	178	0.0095309
21	(D.F.A.B)	172	0.0092097
22	(C.E.F.G)	170	0.0091026
23	(C.F.G)	168	0.0089955
24	(C.D.F)	148	0.0079246
25	(C.F#.A)	145	0.007764
26	(C#.E.A)	132	0.0070679
27	(C.G.A)	124	0.0066395
28	(C.F.G.A)	124	0.0066395
29	(C.E)	113	0.0060505
30	(F.G.B)	112	0.005997
31	(D.G.A#)	111	0.0059435
32	(D.G.A)	110	0.0058899
33	(C#.E.G.A)	109	0.0058364
34	(D.E.G#.B)	107	0.0057293
35	(G.B)	107	0.0057293
36	(C.D.G.A)	87	0.0046584
37	(D.F.A#)	86	0.0046048
38	(C.E.A.B)	85	0.0045513
39	(D.G.A.B)	74	0.0039623
40	(D.E.G)	73	0.0039088
41	(C.D.G.B)	72	0.0038552
42	(D.F.G)	69	0.0036946
43	(C.G)	69	0.0036946
44	(D.E.F.A)	68	0.003641
45	(E.G.A#)	66	0.0035339
46	(C.D.A)	62	0.0033198
47	(C.D.E.A)	59	0.0031591
48	(D.E.A)	59	0.0031591
49	(C.F.A.B)	59	0.0031591
50	(D.F.G.A)	58	0.0031056

50 most frequent pc sets in minor using dense segmentation

		f	*p*
1	(C.D#.G)	2526	0.13879
2	(D#.G.A#)	1458	0.08011
3	(D.G.B)	1317	0.072363
4	(D.F.A#)	1047	0.057527
5	(C.F.G#)	877	0.048187
6	(D.F.G.B)	653	0.035879
7	(C.D#.G#)	621	0.034121
8	(D.G.A#)	559	0.030714
9	(C.D.F.G#)	427	0.023462
10	(C.F.A)	364	0.02
11	(D.F.G#.A#)	356	0.01956
12	(C.E.G)	356	0.01956
13	(D.F.G#)	337	0.018516
14	(C.D#.G.A#)	332	0.018242
15	(C.D#.F.G#)	298	0.016374
16	(C.D.G)	297	0.016319
17	(C.D#.F.A)	274	0.015055
18	(C.D.D#.G)	230	0.012637
19	(D.F.B)	229	0.012582
20	(D.D#.G.A#)	177	0.0097253
21	(D.F.G.A#)	159	0.0087363
22	(C.D#.A)	155	0.0085165
23	(D.F#.A)	145	0.007967
24	(C.F.G)	144	0.0079121
25	(D#.F.A#)	141	0.0077473
26	(C.E.G.A#)	141	0.0077473
27	(C.D.F.G)	139	0.0076374
28	(C.D#.F.G)	138	0.0075824
29	(D.F.G#.B)	136	0.0074725
30	(C.D.F#.A)	135	0.0074176
31	(C.D#.G.G#)	134	0.0073626
32	(D#.F.G.A#)	132	0.0072527
33	(C.D#.G.A)	126	0.0069231
34	(C.D.F)	119	0.0065385
35	(C.F.G.G#)	119	0.0065385
36	(C.D#.F#.A)	104	0.0057143
37	(D.F.A)	98	0.0053846
38	(C.F.A#)	90	0.0049451
39	(C.D#)	90	0.0049451
40	(F.G.B)	73	0.004011
41	(C.D#.F)	73	0.004011
42	(D#.G#.A#)	71	0.0039011
43	(C.G.G#)	68	0.0037363
44	(D#.G)	67	0.0036813
45	(D#.G.B)	67	0.0036813
46	(D.G.A)	60	0.0032967
47	(C.F#.A)	54	0.002967
48	(D.D#.G)	50	0.0027473
49	(C.G.A)	50	0.0027473
50	(C.G)	50	0.0027473

50 most frequent pc sets in major using metrical segmentation

		f	***p***
1	(C.E.G)	3004	0.26024
2	(D.G.B)	1854	0.16062
3	(C.F.A)	910	0.078836
4	(C.E.A)	832	0.072078
5	(D.F.A)	494	0.042797
6	(D.F.G.B)	290	0.025123
7	(E.G.B)	277	0.023997
8	(C.D.G)	259	0.022438
9	(C.D.F.A)	256	0.022178
10	(D.F#.A)	254	0.022005
11	(E.G#.B)	175	0.015161
12	(C.D.F#.A)	164	0.014208
13	(C.E.G.A)	163	0.014121
14	(D.F.B)	140	0.012129
15	(C.D.E.G)	128	0.011089
16	(C#.E.A)	102	0.0088365
17	(C.E)	84	0.0072771
18	(G.B)	78	0.0067573
19	(C.F.G)	76	0.0065841
20	(D.G.A)	72	0.0062375
21	(C.E.G.A#)	71	0.0061509
22	(C.E.F.A)	70	0.0060643
23	(D.F.A#)	67	0.0058044
24	(D.G.A#)	65	0.0056311
25	(C.D.F)	64	0.0055445
26	(C#.E.G.A)	62	0.0053712
27	(C.G.A)	59	0.0051113
28	(C.F.G.A)	57	0.0049381
29	(D.F.A.B)	52	0.0045049
30	(C.F#.A)	49	0.004245
31	(F.G.B)	48	0.0041584
32	(C.D.F.G)	45	0.0038985
33	(C.E.A.B)	44	0.0038118
34	(E.A.B)	43	0.0037252
35	(C.A)	39	0.0033787
36	(D.E.F.A)	38	0.003292
37	(D.G.A.B)	38	0.003292
38	(D.E.G)	37	0.0032054
39	(D.E.G#.B)	36	0.0031188
40	(D.E.G.B)	34	0.0029455
41	(E.G.A#)	31	0.0026856
42	(D.F.G#.B)	31	0.0026856
43	(C.D.G.A)	30	0.002599
44	(C.D.F#)	30	0.002599
45	(C#.E.G)	29	0.0025123
46	(D.E.G.A#)	29	0.0025123
47	(C.E.G.B)	28	0.0024257
48	(D.E.A)	28	0.0024257
49	(C.G)	27	0.0023391
50	(D.A.B)	24	0.0020792

50 most frequent pc sets in minor using metrical segmentation

		f	*p*
1	(C.D#.G)	2075	0.1927
2	(D#.G.A#)	1170	0.10866
3	(D.G.B)	1101	0.10225
4	(D.F.A#)	836	0.077637
5	(C.F.G#)	570	0.052935
6	(C.D#.G#)	488	0.045319
7	(D.G.A#)	404	0.037519
8	(C.D.F.G#)	309	0.028696
9	(C.E.G)	296	0.027489
10	(C.F.A)	256	0.023774
11	(C.D.G)	202	0.018759
12	(C.D#.F.G#)	150	0.01393
13	(D.F.G#)	141	0.013094
14	(D.F.G.B)	141	0.013094
15	(C.D#.F.A)	120	0.011144
16	(D.F#.A)	107	0.0099368
17	(C.D.D#.G)	102	0.0094725
18	(D#.F.A#)	99	0.0091939
19	(D.F.G#.A#)	99	0.0091939
20	(D.F.B)	94	0.0087296
21	(C.D#.G.A#)	80	0.0074294
22	(D.F.G#.B)	78	0.0072437
23	(C.F.G)	77	0.0071508
24	(C.D#.F#.A)	77	0.0071508
25	(D#.F.G.A#)	66	0.0061293
26	(C.D#)	65	0.0060364
27	(C.E.G.A#)	64	0.0059435
28	(C.D#.G.A)	64	0.0059435
29	(C.F.A#)	62	0.0057578
30	(C.F.G.G#)	62	0.0057578
31	(D.F.A)	55	0.0051077
32	(C.D#.G.G#)	53	0.004922
33	(C.G.G#)	53	0.004922
34	(C.D.F)	52	0.0048291
35	(C.D#.A)	51	0.0047363
36	(D#.G)	47	0.0043648
37	(D#.G#.A#)	46	0.0042719
38	(C.D.F#.A)	38	0.003529
39	(C.D.F.G)	29	0.0026932
40	(D#.G.B)	29	0.0026932
41	(C#.F.A#)	28	0.0026003
42	(D.G.A)	26	0.0024146
43	(D.D#.G.A#)	25	0.0023217
44	(C#.F.G.A#)	25	0.0023217
45	(D.F.G.A#)	25	0.0023217
46	(D.G.A.A#)	24	0.0022288
47	(D.D#.G)	23	0.002136
48	(F.G.A#)	22	0.0020431
49	(D#.F.G#)	21	0.0019502
50	(D.F.G.G#)	21	0.0019502

50 most frequent pc sets in major using harmonic approximation

		f	*p*
1	(C.E.G)	3217	0.2787
2	(D.G.B)	2068	0.17916
3	(C.F.A)	990	0.085766
4	(C.E.A)	902	0.078143
5	(D.F.A)	627	0.054319
6	(D.F.G.B)	383	0.03318
7	(D.F#.A)	316	0.027376
8	(E.G.B)	314	0.027203
9	(C.D.F#.A)	210	0.018193
10	(E.G#.B)	201	0.017413
11	(C.D.F.A)	180	0.015594
12	(C.E.G.A)	139	0.012042
13	(D.F.B)	136	0.011782
14	(C#.E.A)	124	0.010742
15	(C.D.G)	106	0.0091831
16	(D.G.A#)	94	0.0081435
17	(C.E.G.A#)	93	0.0080568
18	(C#.E.G.A)	77	0.0066707
19	(D.F.A#)	77	0.0066707
20	(C.E)	77	0.0066707
21	(F.G.B)	74	0.0064108
22	(G.B)	62	0.0053712
23	(C.G.A)	51	0.0044183
24	(C.D.F)	51	0.0044183
25	(D.E.G#.B)	45	0.0038985
26	(C.F#.A)	43	0.0037252
27	(C.D.F#)	42	0.0036386
28	(C.A)	36	0.0031188
29	(C.G)	35	0.0030321
30	(C#.E.G)	33	0.0028589
31	(D.G.A)	31	0.0026856
32	(E.G.A#)	30	0.002599
33	(C.E.F.A)	27	0.0023391
34	(D.F.A.B)	26	0.0022524
35	(C.E.G.B)	23	0.0019925
36	(D.F)	23	0.0019925
37	(D.F#.B)	21	0.0018193
38	(D.F.G#.B)	21	0.0018193
39	(D.E.G)	21	0.0018193
40	(D.E.G.B)	21	0.0018193
41	(D.G#.B)	20	0.0017327
42	(F.A)	19	0.001646
43	(E.G)	19	0.001646
44	(D.E.G.A#)	19	0.001646
45	(D.G)	18	0.0015594
46	(D#.F#.B)	16	0.0013861
47	(C.F.G)	16	0.0013861
48	(E.A)	15	0.0012995
49	(D.F.G)	15	0.0012995
50	(D.A)	15	0.0012995

50 most frequent pc sets in minor using harmonic approximation

		f	***p***
1	(C.D#.G)	2293	0.21295
2	(D#.G.A#)	1305	0.12119
3	(D.G.B)	1249	0.11599
4	(D.F.A#)	958	0.088967
5	(C.F.G#)	769	0.071415
6	(C.D#.G#)	523	0.04857
7	(D.G.A#)	456	0.042348
8	(C.E.G)	336	0.031204
9	(C.F.A)	316	0.029346
10	(D.F.G.B)	222	0.020617
11	(C.D.F.G#)	156	0.014487
12	(D.F.G#)	154	0.014302
13	(C.D#.F.A)	136	0.01263
14	(D.F#.A)	132	0.012259
15	(D.F.G#.A#)	131	0.012166
16	(C.D#.F.G#)	115	0.01068
17	(D.F.B)	98	0.009101
18	(C.E.G.A#)	86	0.0079866
19	(D.F.A)	73	0.0067793
20	(C.D.F#.A)	73	0.0067793
21	(C.D.G)	68	0.006315
22	(C.D#.G.A#)	60	0.0055721
23	(C.D.F)	59	0.0054792
24	(C.D#)	56	0.0052006
25	(C.D#.F#.A)	52	0.0048291
26	(C.D#.A)	48	0.0044577
27	(D#.G)	40	0.0037147
28	(C#.F.A#)	37	0.0034361
29	(F.G.B)	36	0.0033432
30	(C.D#.G.A)	36	0.0033432
31	(D.F.G#.B)	27	0.0025074
32	(D#.F.A#)	27	0.0025074
33	(D.G#.A#)	24	0.0022288
34	(G.A#)	22	0.0020431
35	(D.F.G.A#)	21	0.0019502
36	(D.G)	20	0.0018574
37	(F.G#)	20	0.0018574
38	(C.G.A)	19	0.0017645
39	(E.G.A#)	18	0.0016716
40	(C.D#.G.G#)	18	0.0016716
41	(C.D#.F)	16	0.0014859
42	(C.D.F#)	16	0.0014859
43	(C#.F.G#)	16	0.0014859
44	(C.F#.A)	14	0.0013001
45	(D.A#)	14	0.0013001
46	(C#.E.G.A#)	13	0.0012073
47	(C.F.G)	13	0.0012073
48	(C.G)	13	0.0012073
49	(C.E.A#)	12	0.0011144
50	(D.D#.G)	12	0.0011144

50 most frequent 2-grams in major using dense segmentation

			f	***p***	***rf***	***rp***
1	(D.G.B)	(C.E.G)	605	0.032747	443	0.0232
2	(D.F.G.B)	(C.E.G)	525	0.028417	157	0.008222
3	(C.E.G)	(C.E.G)	480	0.025981	623	0.032626
4	(D.G.B)	(D.F.G.B)	428	0.023166	101	0.0052893
5	(C.E.G)	(D.G.B)	364	0.019702	437	0.022886
6	(C.E.G)	(C.F.A)	292	0.015805	204	0.010683
7	(C.E.G)	(C.E.G.B)	220	0.011908	45	0.0023566
8	(D.F.B)	(C.E.G)	210	0.011367	57	0.0029851
9	(D.G.B)	(D.G.B)	207	0.011204	275	0.014402
10	(C.D.F#.A)	(D.G.B)	186	0.010068	41	0.0021472
11	(C.F.A)	(C.E.G)	177	0.0095805	219	0.011469
12	(C.D.G)	(D.G.B)	176	0.0095264	56	0.0029327
13	(C.E.G)	(C.D.F.A)	169	0.0091475	74	0.0038754
14	(C.E.G)	(C.D.F.G)	150	0.0081191	41	0.0021472
15	(C.D.E.G)	(C.E.G)	144	0.0077943	46	0.002409
16	(C.F.A)	(C.E.F.A)	143	0.0077402	24	0.0012569
17	(C.E.G)	(C.D.G)	139	0.0075237	69	0.0036135
18	(C.E.G)	(C.E.F.G)	129	0.0069824	24	0.0012569
19	(C.D.F.A)	(D.G.B)	125	0.0067659	58	0.0030374
20	(C.D.F.G)	(C.E.G)	119	0.0064411	33	0.0017282
21	(C.E.F.A)	(D.G.B)	118	0.006387	32	0.0016758
22	(D.F#.A)	(C.D.F#.A)	115	0.0062246	8	0.00041896
23	(C.E.G)	(C.E.A)	114	0.0061705	196	0.010264
24	(C.E.G)	(C.D.E.G)	112	0.0060622	43	0.0022519
25	(C.E.A)	(C.E.G.A)	110	0.005954	15	0.00078555
26	(D.F#.A)	(D.G.B)	108	0.0058457	41	0.0021472
27	(C.F#.A)	(D.G.B)	101	0.0054668	16	0.00083792
28	(C.E.G.A#)	(C.F.A)	100	0.0054127	10	0.0005237
29	(C.E.A)	(D.G.B)	100	0.0054127	128	0.0067033
30	(D.G.B)	(C.E.A)	94	0.005088	117	0.0061273
31	(C.E.G)	(C.E.G.A#)	92	0.0049797	31	0.0016235
32	(C.E.G)	(C.D.F)	91	0.0049256	35	0.0018329
33	(C.E.A)	(C.F.A)	90	0.0048714	68	0.0035611
34	(D.F.A)	(D.F.B)	89	0.0048173	13	0.00068081
35	(D.F.A.B)	(C.E.G)	86	0.0046549	31	0.0016235
36	(C.E.A)	(C.E.G)	85	0.0046008	172	0.0090076
37	(C.F.A)	(D.F.A.B)	84	0.0045467	13	0.00068081
38	(F.G.B)	(C.E.G)	84	0.0045467	14	0.00073318
39	(E.G#.B)	(C.E.A)	81	0.0043843	10	0.0005237
40	(C.E.G.B)	(C.F.A)	78	0.0042219	23	0.0012045
41	(C.F.A)	(D.G.B)	76	0.0041137	134	0.0070175
42	(D.G.B)	(E.G.B)	74	0.0040054	70	0.0036659
43	(C.F.G)	(C.E.G)	74	0.0040054	32	0.0016758
44	(C.E.G)	(C.F.G)	72	0.0038972	30	0.0015711
45	(C.E.F.G)	(D.G.B)	69	0.0037348	18	0.00094266
46	(C.E.G.B)	(C.E.A)	66	0.0035724	12	0.00062844
47	(D.G.B)	(C.E.G.A)	66	0.0035724	59	0.0030898
48	(C.E.G)	(C.D.F#.A)	63	0.00341	72	0.0037706
49	(C.E.G)	(D.F.A)	63	0.00341	129	0.0067557
50	(D.F.G.B)	(C.E.A)	61	0.0033018	46	0.002409

50 most frequent 2-grams in minor using dense segmentation

			f	***p***	***rf***	***rp***
1	(D.G.B)	(C.D#.G)	401	0.022259	182	0.010356
2	(D.G.B)	(D.F.G.B)	392	0.02176	44	0.0025036
3	(D.F.G.B)	(C.D#.G)	337	0.018707	74	0.0042105
4	(C.D#.G)	(C.D#.G)	329	0.018263	364	0.020711
5	(D.F.A#)	(D#.G.A#)	256	0.01421	89	0.005064
6	(C.D#.G)	(D.G.B)	237	0.013156	172	0.0097866
7	(C.D#.G)	(C.D.F.G#)	195	0.010824	69	0.003926
8	(D#.G.A#)	(D#.G.A#)	190	0.010547	118	0.0067141
9	(D.F.A#)	(D.F.G#.A#)	189	0.010491	26	0.0014794
10	(C.D#.G)	(C.D#.G.A#)	187	0.01038	48	0.0027312
11	(D.F.G#.A#)	(D#.G.A#)	175	0.0097141	29	0.0016501
12	(D.G.B)	(D.G.B)	141	0.0078268	92	0.0052347
13	(D.F.B)	(C.D#.G)	136	0.0075493	22	0.0012518
14	(C.D.G)	(D.G.B)	135	0.0074938	23	0.0013087
15	(C.D.F.G#)	(D.G.B)	132	0.0073272	32	0.0018208
16	(C.D#.G)	(C.D.G)	129	0.0071607	55	0.0031294
17	(C.D.D#.G)	(C.D#.G)	119	0.0066056	47	0.0026743
18	(C.D#.F.A)	(D.F.A#)	114	0.0063281	19	0.0010811
19	(C.F.G#)	(D.G.B)	113	0.0062726	53	0.0030156
20	(C.D#.G)	(C.D.D#.G)	112	0.006217	26	0.0014794
21	(D.F.G#)	(D#.G.A#)	105	0.0058285	16	0.00091038
22	(C.D#.G)	(C.D#.F.G)	103	0.0057175	21	0.0011949
23	(C.D#.G)	(C.D.F.G)	103	0.0057175	23	0.0013087
24	(D#.G.A#)	(D.F.A#)	98	0.0054399	68	0.0038691
25	(C.D.F.G)	(C.D#.G)	97	0.0053844	19	0.0010811
26	(D#.G.A#)	(C.D#.G#)	96	0.0053289	44	0.0025036
27	(C.F.A)	(C.D#.F.A)	94	0.0052179	3	0.0001707
28	(D.F.A#)	(D.F.A#)	91	0.0050513	64	0.0036415
29	(C.F.A)	(D.F.A#)	91	0.0050513	21	0.0011949
30	(C.D#.G)	(D.F.A#)	88	0.0048848	163	0.0092745
31	(D#.G.A#)	(D.D#.G.A#)	85	0.0047183	18	0.0010242
32	(D.F.G.B)	(C.E.G)	82	0.0045518	9	0.00051209
33	(D#.G.A#)	(C.D#.G)	82	0.0045518	200	0.01138
34	(D#.F.A#)	(D.F.A#)	81	0.0044963	9	0.00051209
35	(D#.G.A#)	(C.D#.F.G#)	80	0.0044407	26	0.0014794
36	(D#.F.G.A#)	(D#.G.A#)	77	0.0042742	7	0.00039829
37	(C.D#.G)	(C.D#.G#)	73	0.0040522	86	0.0048933
38	(C.D#.G)	(C.F.G#)	73	0.0040522	102	0.0058037
39	(C.D#.A)	(D.F.A#)	70	0.0038857	4	0.0002276
40	(C.D#.G)	(D#.G.A#)	69	0.0038301	206	0.011721
41	(C.D#.G#)	(D#.G.A#)	69	0.0038301	48	0.0027312
42	(C.D#.G)	(C.D.F)	68	0.0037746	12	0.00068279
43	(D.F.G.B)	(C.D#.G#)	68	0.0037746	26	0.0014794
44	(C.F.G#)	(C.D#.G)	65	0.0036081	124	0.0070555
45	(D#.G.A#)	(D#.F.A#)	64	0.0035526	7	0.00039829
46	(C.D#.G)	(D.G.A#)	63	0.0034971	76	0.0043243
47	(D#.G.A#)	(D#.F.G.A#)	63	0.0034971	14	0.00079659
48	(D.G.A#)	(D.F.G.A#)	63	0.0034971	5	0.0002845
49	(C.D#.G)	(C.D#.A)	60	0.0033306	16	0.00091038
50	(D.F.G#.B)	(C.D#.G)	60	0.0033306	23	0.0013087

50 most frequent 2-grams in major using metrical segmentation

			f	***p***	***rf***	***rp***
1	(D.G.B)	(C.E.G)	833	0.073444	459	0.040063
2	(C.E.G)	(C.E.G)	690	0.060836	804	0.070175
3	(C.E.G)	(D.G.B)	539	0.047522	479	0.041809
4	(C.E.G)	(C.F.A)	411	0.036237	235	0.020511
5	(C.F.A)	(C.E.G)	280	0.024687	228	0.0199
6	(D.G.B)	(D.G.B)	228	0.020102	286	0.024963
7	(D.F.G.B)	(C.E.G)	204	0.017986	86	0.0075063
8	(C.E.G)	(C.E.A)	182	0.016047	242	0.021122
9	(D.F#.A)	(D.G.B)	156	0.013754	37	0.0032295
10	(C.D.F.A)	(D.G.B)	154	0.013578	31	0.0027058
11	(C.E.G)	(C.D.G)	132	0.011638	56	0.0048878
12	(C.E.A)	(D.G.B)	117	0.010316	132	0.011521
13	(C.F.A)	(D.G.B)	108	0.0095221	142	0.012394
14	(D.G.B)	(C.E.A)	106	0.0093458	128	0.011172
15	(C.E.G)	(C.D.F.A)	104	0.0091695	62	0.0054115
16	(C.D.G)	(D.G.B)	94	0.0082878	42	0.0036659
17	(E.G#.B)	(C.E.A)	94	0.0082878	16	0.0013965
18	(D.F.A)	(C.E.G)	92	0.0081114	108	0.0094266
19	(C.D.G)	(C.E.G)	89	0.0078469	63	0.0054988
20	(C.E.A)	(C.E.A)	88	0.0077588	54	0.0047133
21	(C.D.F#.A)	(D.G.B)	84	0.0074061	22	0.0019202
22	(C.E.A)	(C.E.G)	83	0.0073179	214	0.018679
23	(D.F.B)	(C.E.G)	71	0.0062599	41	0.0035786
24	(C.E.G)	(D.F.A)	70	0.0061718	136	0.01187
25	(C.F.A)	(C.F.A)	70	0.0061718	82	0.0071572
26	(C.E.A)	(C.D.F.A)	66	0.0058191	18	0.0015711
27	(D.G.B)	(C.F.A)	63	0.0055546	137	0.011958
28	(D.G.B)	(E.G.B)	56	0.0049374	48	0.0041896
29	(C.E.G.A)	(D.F#.A)	55	0.0048492	1	8.7283e-005
30	(D.F.A)	(D.G.B)	54	0.0047611	75	0.0065462
31	(C.E.A)	(E.G.B)	53	0.0046729	20	0.0017457
32	(E.G.B)	(C.F.A)	52	0.0045847	17	0.0014838
33	(D.G.B)	(D.F.A)	51	0.0044966	69	0.0060225
34	(C.F.A)	(D.F.A)	50	0.0044084	33	0.0028803
35	(C.E.A)	(C.F.A)	50	0.0044084	50	0.0043641
36	(D.F.A)	(C.E.A)	50	0.0044084	33	0.0028803
37	(D.G.B)	(D.F.G.B)	49	0.0043202	49	0.0042769
38	(C.E.A)	(D.F.A)	48	0.0042321	35	0.0030549
39	(D.G.B)	(D.F#.A)	47	0.0041439	43	0.0037532
40	(C.E.G)	(D.F.G.B)	45	0.0039676	71	0.0061971
41	(C.E.G)	(C.D.F#.A)	44	0.0038794	40	0.0034913
42	(C.E.A)	(C.D.F#.A)	43	0.0037912	9	0.00078555
43	(D.G.B)	(C.E.G.A)	42	0.0037031	35	0.0030549
44	(C.E.G)	(D.F.B)	42	0.0037031	38	0.0033167
45	(C.F.A)	(D.F.B)	41	0.0036149	8	0.00069826
46	(D.F.A)	(D.F.A)	41	0.0036149	20	0.0017457
47	(C#.E.A)	(D.F.A)	41	0.0036149	3	0.00026185
48	(E.G.B)	(C.E.A)	39	0.0034385	14	0.001222
49	(D.G.B)	(C.D.E.G)	39	0.0034385	31	0.0027058
50	(C.D.E.G)	(C.D.G)	38	0.0033504	5	0.00043641

50 most frequent 2-grams in minor using metrical segmentation

			f	***p***	***rf***	***rp***
1	(D.G.B)	(C.D#.G)	552	0.052159	227	0.021156
2	(C.D#.G)	(C.D#.G)	500	0.047246	398	0.037092
3	(D.F.A#)	(D#.G.A#)	346	0.032694	91	0.0084809
4	(C.D#.G)	(D.G.B)	298	0.028158	190	0.017707
5	(D#.G.A#)	(D#.G.A#)	263	0.024851	128	0.011929
6	(C.D.F.G#)	(D.G.B)	170	0.016063	35	0.0032619
7	(D#.G.A#)	(D.F.A#)	163	0.015402	89	0.0082945
8	(C.D#.G)	(C.D.F.G#)	137	0.012945	62	0.0057782
9	(D#.G.A#)	(C.D#.G#)	134	0.012662	56	0.005219
10	(D.G.B)	(D.G.B)	133	0.012567	110	0.010252
11	(C.D#.G)	(C.F.G#)	132	0.012473	115	0.010718
12	(C.D#.G)	(D.F.A#)	116	0.010961	156	0.014539
13	(C.D#.G)	(C.D.G)	115	0.010866	39	0.0036347
14	(D#.G.A#)	(C.D#.G)	114	0.010772	222	0.02069
15	(C.F.A)	(D.F.A#)	107	0.010111	16	0.0014911
16	(C.D#.G#)	(D#.G.A#)	107	0.010111	61	0.005685
17	(D.G.B)	(C.E.G)	101	0.0095436	34	0.0031687
18	(D.F.A#)	(D.F.A#)	91	0.0085987	56	0.005219
19	(C.F.G#)	(C.D#.G)	91	0.0085987	114	0.010624
20	(D.F.G.B)	(C.D#.G)	79	0.0074648	26	0.0024231
21	(C.D#.G)	(D#.G.A#)	74	0.0069923	216	0.02013
22	(C.D#.G)	(D.G.A#)	73	0.0068979	54	0.0050326
23	(C.F.G#)	(C.F.G#)	73	0.0068979	20	0.0018639
24	(C.D.G)	(D.G.B)	72	0.0068034	18	0.0016775
25	(C.D#.G)	(C.D#.G#)	72	0.0068034	99	0.0092265
26	(C.D#.F.G#)	(D.F.A#)	63	0.0059529	13	0.0012116
27	(D.G.A#)	(D.G.A#)	63	0.0059529	8	0.00074557
28	(C.D#.G)	(C.F.A)	61	0.005764	54	0.0050326
29	(D.F.B)	(C.D#.G)	59	0.005575	18	0.0016775
30	(C.D.G)	(C.D#.G)	58	0.0054805	35	0.0032619
31	(D.G.B)	(C.D#.G#)	57	0.005386	54	0.0050326
32	(C.E.G)	(C.F.G#)	55	0.005197	20	0.0018639
33	(C.D#.F.A)	(D.F.A#)	53	0.005008	14	0.0013048
34	(D.F.A#)	(C.D#.G)	50	0.0047246	168	0.015657
35	(C.F.G#)	(D.G.B)	50	0.0047246	57	0.0053122
36	(C.D#.G)	(C.D#.F.G#)	49	0.0046301	31	0.0028891
37	(D.F.G#.A#)	(D#.G.A#)	47	0.0044411	13	0.0012116
38	(C.D#.G#)	(C.D.F.G#)	46	0.0043466	12	0.0011184
39	(D#.G.A#)	(D#.F.A#)	45	0.0042521	10	0.00093197
40	(C.D#.G)	(C.D#.F.A)	45	0.0042521	21	0.0019571
41	(D.F#.A)	(D.G.A#)	45	0.0042521	4	0.00037279
42	(D#.G.A#)	(C.F.G#)	44	0.0041576	55	0.0051258
43	(D#.G.A#)	(C.D#.F.G#)	44	0.0041576	14	0.0013048
44	(C.D.F.G#)	(C.D#.G)	43	0.0040631	44	0.0041007
45	(D.G.A#)	(C.D#.G)	43	0.0040631	71	0.006617
46	(D.G.A#)	(D#.G.A#)	41	0.0038741	41	0.0038211
47	(D.F.A#)	(C.F.G#)	41	0.0038741	40	0.0037279
48	(C.D#.G#)	(C.D#.G#)	39	0.0036852	15	0.0013979
49	(C.D#.G#)	(C.F.G#)	39	0.0036852	30	0.0027959
50	(C.D#.G)	(D.F.G#)	38	0.0035907	35	0.0032619

50 most frequent 2-grams in major using harmonic approximation

			f	***p***	***rf***	***rp***
1	(D.G.B)	(C.E.G)	1019	0.089843	579	0.050537
2	(C.E.G)	(C.E.G)	723	0.063745	894	0.078031
3	(C.E.G)	(D.G.B)	657	0.057926	583	0.050886
4	(C.E.G)	(C.F.A)	468	0.041263	270	0.023566
5	(C.F.A)	(C.E.G)	345	0.030418	266	0.023217
6	(D.F.G.B)	(C.E.G)	290	0.025569	107	0.0093393
7	(D.G.B)	(D.G.B)	256	0.022571	386	0.033691
8	(D.F#.A)	(D.G.B)	192	0.016928	46	0.004015
9	(C.E.G)	(C.E.A)	191	0.01684	267	0.023305
10	(C.E.A)	(D.G.B)	156	0.013754	171	0.014925
11	(D.G.B)	(C.E.A)	145	0.012784	153	0.013354
12	(C.F.A)	(D.G.B)	138	0.012167	164	0.014314
13	(C.E.G)	(D.F.A)	130	0.011462	154	0.013442
14	(C.D.F.A)	(D.G.B)	127	0.011197	26	0.0022694
15	(E.G#.B)	(C.E.A)	120	0.01058	13	0.0011347
16	(C.D.F#.A)	(D.G.B)	118	0.010404	35	0.0030549
17	(D.F.A)	(C.E.G)	112	0.0098748	177	0.015449
18	(C.E.A)	(C.E.G)	104	0.0091695	233	0.020337
19	(C.E.G)	(C.D.F.A)	100	0.0088168	47	0.0041023
20	(C.E.A)	(C.E.A)	96	0.0084641	68	0.0059352
21	(C.E.G)	(D.F.G.B)	95	0.0083759	123	0.010736
22	(D.F.A)	(D.G.B)	92	0.0081114	110	0.0096011
23	(D.F.B)	(C.E.G)	91	0.0080233	43	0.0037532
24	(C.F.A)	(C.F.A)	79	0.0069653	66	0.0057607
25	(C.D.G)	(D.G.B)	76	0.0067008	20	0.0017457
26	(D.G.B)	(C.F.A)	72	0.0063481	156	0.013616
27	(E.G.B)	(C.F.A)	72	0.0063481	32	0.0027931
28	(C.E.A)	(D.F.A)	71	0.0062599	52	0.0045387
29	(C.E.G)	(C.D.G)	71	0.0062599	25	0.0021821
30	(D.G.B)	(D.F#.A)	64	0.0056427	52	0.0045387
31	(C.F.A)	(D.F.A)	63	0.0055546	66	0.0057607
32	(C.E.A)	(C.F.A)	63	0.0055546	73	0.0063717
33	(C.E.G.A#)	(C.F.A)	63	0.0055546	11	0.00096011
34	(D.G.B)	(E.G.B)	62	0.0054664	60	0.005237
35	(C.E.A)	(E.G.B)	62	0.0054664	34	0.0029676
36	(C.E.G)	(C.D.F#.A)	61	0.0053782	68	0.0059352
37	(D.G.B)	(D.F.A)	58	0.0051137	108	0.0094266
38	(D.G.B)	(D.F.G.B)	58	0.0051137	59	0.0051497
39	(C.F.A)	(D.F.G.B)	57	0.0050256	32	0.0027931
40	(D.F.A)	(C.E.A)	57	0.0050256	53	0.004626
41	(C.E.G)	(D.F#.A)	55	0.0048492	87	0.0075936
42	(F.G.B)	(C.E.G)	55	0.0048492	28	0.0024439
43	(C.E.G.A)	(D.F#.A)	53	0.0046729	8	0.00069826
44	(C#.E.A)	(D.F.A)	52	0.0045847	3	0.00026185
45	(D.F.A)	(D.F.G.B)	50	0.0044084	23	0.0020075
46	(C.E.A)	(C.D.F#.A)	50	0.0044084	18	0.0015711
47	(C.E.A)	(C.D.F.A)	49	0.0043202	10	0.00087283
48	(D.F.A)	(D.F.A)	49	0.0043202	34	0.0029676
49	(E.G.B)	(C.E.A)	48	0.0042321	22	0.0019202
50	(C#.E.G.A)	(D.F.A)	46	0.0040557	2	0.00017457

50 most frequent 2-grams in minor using harmonic approximation

			f	*p*	*rf*	*rp*
1	(D.G.B)	(C.D#.G)	660	0.062364	269	0.02507
2	(C.D#.G)	(C.D#.G)	544	0.051403	532	0.049581
3	(D.F.A#)	(D#.G.A#)	435	0.041104	108	0.010065
4	(C.D#.G)	(D.G.B)	414	0.039119	266	0.02479
5	(D#.G.A#)	(D#.G.A#)	282	0.026647	155	0.014445
6	(D#.G.A#)	(D.F.A#)	224	0.021166	101	0.0094129
7	(C.D#.G)	(C.F.G#)	187	0.01767	185	0.017241
8	(D#.G.A#)	(C.D#.G#)	176	0.01663	57	0.0053122
9	(C.F.A)	(D.F.A#)	156	0.014741	26	0.0024231
10	(D.G.B)	(D.G.B)	151	0.014268	142	0.013234
11	(C.D#.G#)	(D#.G.A#)	151	0.014268	45	0.0041938
12	(C.D#.G)	(D.F.A#)	149	0.014079	212	0.019758
13	(D.F.G.B)	(C.D#.G)	141	0.013323	53	0.0049394
14	(D#.G.A#)	(C.D#.G)	132	0.012473	247	0.02302
15	(C.F.G#)	(D.G.B)	129	0.012189	107	0.009972
16	(C.F.G#)	(C.D#.G)	124	0.011717	168	0.015657
17	(D.G.B)	(C.E.G)	117	0.011055	43	0.0040075
18	(C.D.F.G#)	(D.G.B)	105	0.0099216	24	0.0022367
19	(D.F.A#)	(D.F.A#)	101	0.0095436	84	0.0078285
20	(C.D#.G)	(C.D.F.G#)	96	0.0090712	34	0.0031687
21	(C.D#.G)	(D.G.A#)	95	0.0089767	96	0.0089469
22	(C.F.G#)	(C.F.G#)	91	0.0085987	56	0.005219
23	(C.E.G)	(C.F.G#)	87	0.0082207	22	0.0020503
24	(C.D#.G)	(D#.G.A#)	85	0.0080317	284	0.026468
25	(C.D#.G)	(C.D#.G#)	78	0.0073703	100	0.0093197
26	(D.F.A#)	(C.D#.G)	78	0.0073703	229	0.021342
27	(D.G.B)	(C.D#.G#)	76	0.0071813	45	0.0041938
28	(D.F.G#.A#)	(D#.G.A#)	74	0.0069923	15	0.0013979
29	(D.F.B)	(C.D#.G)	73	0.0068979	22	0.0020503
30	(D#.G.A#)	(C.F.G#)	71	0.0067089	90	0.0083877
31	(C.D#.G)	(C.F.A)	70	0.0066144	81	0.0075489
32	(D.G.A#)	(D.G.A#)	70	0.0066144	24	0.0022367
33	(C.D#.G#)	(C.F.G#)	61	0.005764	39	0.0036347
34	(C.D#.F.A)	(D.F.A#)	60	0.0056695	13	0.0012116
35	(C.D#.G)	(D.F.G.B)	59	0.005575	51	0.004753
36	(D.F#.A)	(D.G.A#)	58	0.0054805	5	0.00046598
37	(C.F.G#)	(C.E.G)	57	0.005386	29	0.0027027
38	(D.G.A#)	(C.D#.G)	56	0.0052915	103	0.0095993
39	(C.D#.F.G#)	(D.F.A#)	55	0.005197	12	0.0011184
40	(C.D.G)	(D.G.B)	54	0.0051025	14	0.0013048
41	(C.F.G#)	(D#.G.A#)	51	0.004819	77	0.0071761
42	(C.D#.G)	(C.D#.F.A)	50	0.0047246	22	0.0020503
43	(C.D#.G)	(C.D.G)	49	0.0046301	20	0.0018639
44	(D.G.A#)	(D#.G.A#)	49	0.0046301	59	0.0054986
45	(D.F.A#)	(C.F.G#)	49	0.0046301	71	0.006617
46	(D.F.A#)	(C.F.A)	46	0.0043466	27	0.0025163
47	(D#.G.A#)	(C.D#.F.G#)	44	0.0041576	14	0.0013048
48	(D.F.G#)	(D.G.B)	44	0.0041576	17	0.0015843
49	(D.F#.A)	(D.G.B)	43	0.0040631	18	0.0016775
50	(D#.G.A#)	(C.F.A)	42	0.0039686	40	0.0037279

50 most frequent 2-grams in major within a random corpus

			f	*p*
1	(C.E.G)	(C.E.G)	623	0.032626
2	(D.G.B)	(C.E.G)	443	0.0232
3	(C.E.G)	(D.G.B)	437	0.022886
4	(D.G.B)	(D.G.B)	275	0.014402
5	(C.F.A)	(C.E.G)	219	0.011469
6	(C.E.G)	(C.F.A)	204	0.010683
7	(C.E.G)	(C.E.A)	196	0.010264
8	(C.E.A)	(C.E.G)	172	0.0090076
9	(C.E.G)	(D.F.G.B)	162	0.0084839
10	(D.F.G.B)	(C.E.G)	157	0.008222
11	(C.F.A)	(D.G.B)	134	0.0070175
12	(C.E.G)	(D.F.A)	129	0.0067557
13	(C.E.A)	(D.G.B)	128	0.0067033
14	(D.G.B)	(C.F.A)	122	0.0063891
15	(D.G.B)	(C.E.A)	117	0.0061273
16	(D.F.G.B)	(D.G.B)	109	0.0057083
17	(D.F.A)	(C.E.G)	108	0.0056559
18	(D.G.B)	(D.F.G.B)	101	0.0052893
19	(E.G.B)	(C.E.G)	99	0.0051846
20	(C.E.G)	(E.G.B)	84	0.0043991
21	(C.E.G)	(D.F.B)	84	0.0043991
22	(C.D.F#.A)	(C.E.G)	80	0.0041896
23	(C.D.G)	(C.E.G)	78	0.0040848
24	(D.G.B)	(D.F.A)	78	0.0040848
25	(C.E.G)	(C.D.F.A)	74	0.0038754
26	(D.F.A)	(D.G.B)	74	0.0038754
27	(C.E.G)	(C.D.F#.A)	72	0.0037706
28	(C.F.A)	(C.E.A)	71	0.0037183
29	(D.G.B)	(E.G.B)	70	0.0036659
30	(C.E.G)	(C.D.G)	69	0.0036135
31	(C.E.G)	(D.F#.A)	69	0.0036135
32	(C.E.A)	(C.F.A)	68	0.0035611
33	(C.F.A)	(C.F.A)	68	0.0035611
34	(D.F#.A)	(C.E.G)	68	0.0035611
35	(C.E.G.A)	(C.E.G)	66	0.0034564
36	(C.D.F.A)	(C.E.G)	66	0.0034564
37	(E.G.B)	(D.G.B)	66	0.0034564
38	(D.F.G.B)	(C.F.A)	64	0.0033517
39	(C.E.F.A)	(C.E.G)	63	0.0032993
40	(C.E.G)	(C.E.G.A)	60	0.0031422
41	(D.G.B)	(C.E.G.A)	59	0.0030898
42	(C.D.F.A)	(D.G.B)	58	0.0030374
43	(C.E.A)	(C.E.A)	58	0.0030374
44	(D.F.B)	(C.E.G)	57	0.0029851
45	(C.D.G)	(D.G.B)	56	0.0029327
46	(C.E.A)	(D.F.G.B)	52	0.0027232
47	(C.F.A)	(D.F.G.B)	50	0.0026185
48	(D.G.B)	(C.D.F.A)	50	0.0026185
49	(C.E.G.B)	(C.E.G)	49	0.0025661
50	(D.F.B)	(D.G.B)	48	0.0025137

50 most frequent 2-grams in minor within a random corpus

			f	***p***
1	(C.D#.G)	(C.D#.G)	364	0.020711
2	(C.D#.G)	(D#.G.A#)	206	0.011721
3	(D#.G.A#)	(C.D#.G)	200	0.01138
4	(D.G.B)	(C.D#.G)	182	0.010356
5	(C.D#.G)	(D.G.B)	172	0.0097866
6	(D.F.A#)	(C.D#.G)	172	0.0097866
7	(C.D#.G)	(D.F.A#)	163	0.0092745
8	(C.F.G#)	(C.D#.G)	124	0.0070555
9	(D#.G.A#)	(D#.G.A#)	118	0.0067141
10	(D#.G.A#)	(D.G.B)	116	0.0066003
11	(D.G.B)	(D.F.A#)	105	0.0059744
12	(C.D#.G)	(C.F.G#)	102	0.0058037
13	(D.G.B)	(D#.G.A#)	102	0.0058037
14	(D.G.B)	(D.G.B)	92	0.0052347
15	(D.F.A#)	(D#.G.A#)	89	0.005064
16	(C.D#.G)	(C.D#.G#)	86	0.0048933
17	(C.D#.G#)	(C.D#.G)	83	0.0047226
18	(C.D#.G)	(D.F.G.B)	79	0.004495
19	(C.D#.G)	(D.G.A#)	76	0.0043243
20	(D.F.G.B)	(C.D#.G)	74	0.0042105
21	(D#.G.A#)	(C.F.G#)	72	0.0040967
22	(C.F.G#)	(D#.G.A#)	72	0.0040967
23	(C.D#.G)	(C.D.F.G#)	69	0.003926
24	(D#.G.A#)	(D.F.A#)	68	0.0038691
25	(D.F.A#)	(D.G.B)	66	0.0037553
26	(D.F.A#)	(D.F.A#)	64	0.0036415
27	(D.G.A#)	(C.D#.G)	64	0.0036415
28	(D.F.A#)	(C.F.G#)	60	0.0034139
29	(D.G.B)	(C.F.G#)	59	0.003357
30	(C.D#.G)	(C.F.A)	59	0.003357
31	(C.F.G#)	(D.F.A#)	56	0.0031863
32	(C.D#.G)	(C.D.G)	55	0.0031294
33	(D#.G.A#)	(D.F.G.B)	54	0.0030725
34	(C.F.G#)	(D.G.B)	53	0.0030156
35	(C.D#.G)	(D.F.G#)	50	0.002845
36	(D.F.G#)	(C.D#.G)	49	0.0027881
37	(C.F.A)	(C.D#.G)	49	0.0027881
38	(C.D#.G.A#)	(C.D#.G)	48	0.0027312
39	(C.D#.G)	(C.D#.G.A#)	48	0.0027312
40	(C.D#.G#)	(D#.G.A#)	48	0.0027312
41	(D.G.B)	(C.D#.G#)	48	0.0027312
42	(D.F.G.B)	(D#.G.A#)	47	0.0026743
43	(C.D.D#.G)	(C.D#.G)	47	0.0026743
44	(C.D.G)	(C.D#.G)	47	0.0026743
45	(C.D.F.G#)	(D#.G.A#)	46	0.0026174
46	(C.D#.G)	(C.D#.F.G#)	46	0.0026174
47	(D.F.G#.A#)	(C.D#.G)	46	0.0026174
48	(C.E.G)	(C.D#.G)	45	0.0025605
49	(C.D.F.G#)	(C.D#.G)	45	0.0025605
50	(D.G.B)	(D.F.G.B)	44	0.0025036

50 most frequent 3-grams in major using dense segmentation

				f	***p***	***rf***	***rp***
1	(D.G.B)	(D.F.G.B)	(C.E.G)	316	0.017292	20	0.0010585
2	(D.F.G.B)	(C.E.G)	(C.E.G)	124	0.0067856	30	0.0015878
3	(C.E.G)	(C.D.F.G)	(C.E.G)	112	0.0061289	8	0.00042341
4	(D.G.B)	(C.E.G)	(C.E.G)	108	0.00591	88	0.0046576
5	(C.E.G)	(D.G.B)	(C.E.G)	87	0.0047609	88	0.0046576
6	(C.E.G)	(C.F.A)	(C.E.F.A)	83	0.004542	4	0.00021171
7	(D.F#.A)	(C.D.F#.A)	(D.G.B)	83	0.004542	2	0.00010585
8	(D.G.B)	(D.G.B)	(C.E.G)	83	0.004542	53	0.0028051
9	(C.E.G)	(C.E.G)	(D.G.B)	79	0.0043231	83	0.0043929
10	(D.G.B)	(C.E.G)	(C.E.G.B)	77	0.0042136	6	0.00031756
11	(C.E.G)	(C.E.G)	(C.F.A)	75	0.0041042	38	0.0020112
12	(C.E.G)	(C.D.F.A)	(D.G.B)	74	0.0040495	11	0.0005822
13	(C.F.A)	(D.F.A.B)	(C.E.G)	70	0.0038306	4	0.00021171
14	(C.D.G)	(D.G.B)	(C.E.G)	70	0.0038306	5	0.00026463
15	(C.E.G)	(C.E.F.G)	(D.G.B)	67	0.0036664	4	0.00021171
16	(D.F.A)	(D.F.B)	(C.E.G)	66	0.0036117	3	0.00015878
17	(C.E.G)	(C.E.G.B)	(C.F.A)	65	0.003557	7	0.00037049
18	(C.E.G)	(C.D.G)	(D.G.B)	63	0.0034475	7	0.00037049
19	(C.E.G)	(C.E.G.B)	(C.E.A)	62	0.0033928	1	5.2927e-005
20	(C.E.G)	(D.G.B)	(D.F.G.B)	61	0.0033381	22	0.0011644
21	(D.F.G.B)	(C.E.G)	(D.G.B)	61	0.0033381	20	0.0010585
22	(C.F.A)	(C.E.F.A)	(D.G.B)	60	0.0032834	2	0.00010585
23	(D.G.B)	(C.E.G)	(D.G.B)	59	0.0032286	52	0.0027522
24	(D.G.B)	(C.E.G)	(C.F.A)	56	0.0030645	23	0.0012173
25	(C.E.G)	(C.D.E.G)	(C.E.G)	55	0.0030097	7	0.00037049
26	(C.E.G)	(C.E.G.A#)	(C.F.A)	53	0.0029003	1	5.2927e-005
27	(D.G.B)	(D.G.B)	(D.F.G.B)	52	0.0028456	8	0.00042341
28	(C.E.G)	(C.E.G)	(C.E.G)	52	0.0028456	107	0.0056632
29	(C.D.F.A)	(D.G.B)	(D.F.G.B)	48	0.0026267	3	0.00015878
30	(C.E.A)	(D.G.B)	(C.E.G)	47	0.002572	19	0.0010056
31	(C.E.G)	(C.D.F)	(C.E.G)	45	0.0024625	7	0.00037049
32	(C.D.G)	(D.G.B)	(D.F.G.B)	44	0.0024078	6	0.00031756
33	(C.F.A)	(C.F.A.B)	(C.E.G)	44	0.0024078	0	0
34	(D.F#.A)	(D.G.B)	(C.E.G)	43	0.0023531	4	0.00021171
35	(C.D.F.A)	(D.G.B)	(C.E.G)	42	0.0022983	12	0.00063512
36	(C.E.G)	(C.E.G)	(C.E.G.B)	41	0.0022436	7	0.00037049
37	(C.D.F#.A)	(D.G.B)	(C.E.G)	40	0.0021889	13	0.00068805
38	(C.E.G)	(C.E.F.G)	(C.D.G)	38	0.0020795	1	5.2927e-005
39	(D.F.B)	(C.E.G)	(C.E.F.G)	37	0.0020247	0	0
40	(D.F.G.B)	(C.D.E.G)	(C.E.G)	37	0.0020247	2	0.00010585
41	(C.D.E.G)	(C.E.G)	(C.D.G)	35	0.0019153	0	0
42	(C.E.G)	(C.E.F.A)	(D.G.B)	35	0.0019153	3	0.00015878
43	(C.E.G)	(D.G.B)	(C.E.A)	33	0.0018058	26	0.0013761
44	(D.G.B)	(D.F.G.B)	(C.E.A)	33	0.0018058	4	0.00021171
45	(C.E.G)	(D.G.B)	(D.G.B)	32	0.0017511	50	0.0026463
46	(C.D.F)	(D.F.B)	(C.E.G)	32	0.0017511	1	5.2927e-005
47	(C.E.G)	(C.F.A)	(C.E.G)	31	0.0016964	41	0.00217
48	(C.E.G)	(D.F.G.B)	(C.E.G)	31	0.0016964	33	0.0017466
49	(D.G.B)	(C.D.G.A)	(D.G.B)	30	0.0016417	3	0.00015878
50	(C.E.G)	(C.E.G)	(C.D.F.G)	30	0.0016417	12	0.00063512

50 most frequent 3-grams in minor using dense segmentation

				f	***p***	***rf***	***rp***
1	(D.G.B)	(D.F.G.B)	(C.D#.G)	227	0.012731	11	0.00063255
2	(D.F.G.B)	(C.D#.G)	(C.D#.G)	115	0.0064498	9	0.00051754
3	(D.F.A#)	(D.F.G#.A#)	(D#.G.A#)	110	0.0061694	4	0.00023002
4	(C.D#.G)	(D.G.B)	(C.D#.G)	99	0.0055524	25	0.0014376
5	(C.D#.G)	(C.D.F.G)	(C.D#.G)	89	0.0049916	3	0.00017251
6	(C.D.F.G#)	(D.G.B)	(D.F.G.B)	83	0.0046551	0	0
7	(C.D#.G)	(C.D.F.G#)	(D.G.B)	80	0.0044868	2	0.00011501
8	(D.G.B)	(D.G.B)	(C.D#.G)	68	0.0038138	15	0.00086256
9	(D.G.B)	(D.F.G.B)	(C.E.G)	67	0.0037577	1	5.7504e-005
10	(C.D#.G)	(D.G.B)	(D.F.G.B)	67	0.0037577	7	0.00040253
11	(C.F.A)	(C.D#.F.A)	(D.F.A#)	61	0.0034212	1	5.7504e-005
12	(D.G.B)	(C.D#.G)	(C.D#.G)	59	0.003309	22	0.0012651
13	(D.F.A#)	(D#.G.A#)	(D#.G.A#)	58	0.0032529	4	0.00023002
14	(D.G.B)	(D.F.G.B)	(C.D#.G#)	49	0.0027482	3	0.00017251
15	(C.D#.G)	(C.D.D#.G)	(C.D#.G)	49	0.0027482	5	0.00028752
16	(C.D.G)	(D.G.B)	(C.D#.G)	48	0.0026921	4	0.00023002
17	(D.G.B)	(C.D#.G)	(C.D#.G.A#)	47	0.002636	2	0.00011501
18	(C.D#.G)	(C.D#.G)	(C.D#.G.A#)	44	0.0024678	7	0.00040253
19	(C.D#.G)	(C.D#.G.A#)	(C.F.G#)	44	0.0024678	2	0.00011501
20	(D.F.G#.A#)	(D#.G.A#)	(D#.G.A#)	42	0.0023556	1	5.7504e-005
21	(C.D.G)	(D.G.B)	(D.F.G.B)	39	0.0021873	1	5.7504e-005
22	(C.D#.G)	(C.D#.F.G)	(D.G.B)	39	0.0021873	2	0.00011501
23	(C.F.G#)	(D.G.B)	(C.D#.G)	39	0.0021873	7	0.00040253
24	(D#.G.A#)	(D#.F.G#.A#)	(D#.G.A#)	38	0.0021312	1	5.7504e-005
25	(C.D#.G)	(C.D#.F.G)	(C.D.G)	37	0.0020752	1	5.7504e-005
26	(C.D#.G)	(C.D.G)	(D.G.B)	36	0.0020191	5	0.00028752
27	(C.D#.G)	(C.D.G)	(C.D#.G)	35	0.001963	12	0.00069005
28	(D#.G.A#)	(D#.G.A#)	(D.F.A#)	35	0.001963	7	0.00040253
29	(D#.G.A#)	(D#.F.G.A#)	(D#.G.A#)	35	0.001963	1	5.7504e-005
30	(C.D.F.G#)	(C.F.G#)	(D.G.B)	34	0.0019069	2	0.00011501
31	(D.F.G.B)	(C.D#.G)	(D.G.B)	34	0.0019069	2	0.00011501
32	(D#.G.A#)	(D#.F.A#)	(D.F.A#)	33	0.0018508	2	0.00011501
33	(C.D#.G)	(C.D#.G.A#)	(C.D#.G#)	33	0.0018508	4	0.00023002
34	(C.D#.F.A)	(D.F.A#)	(D#.G.A#)	32	0.0017947	3	0.00017251
35	(C.D#.G)	(C.D#.G)	(C.D.F.G#)	32	0.0017947	17	0.00097757
36	(C.D#.G)	(C.D#.G)	(D.G.B)	32	0.0017947	25	0.0014376
37	(D.G.B)	(C.D#.G)	(D.G.B)	31	0.0017386	5	0.00028752
38	(D#.F.A#)	(D.F.A#)	(D#.G.A#)	31	0.0017386	1	5.7504e-005
39	(C.D#.G)	(D.F.A#)	(D#.G.A#)	31	0.0017386	5	0.00028752
40	(D.F.A#)	(D#.G.A#)	(D.D#.G.A#)	30	0.0016826	3	0.00017251
41	(C.D#.G)	(D.F.B)	(C.D#.G)	30	0.0016826	2	0.00011501
42	(D.G.B)	(D.G.B)	(D.F.G.B)	29	0.0016265	4	0.00023002
43	(C.D#.G)	(C.D#.A)	(D.F.A#)	29	0.0016265	0	0
44	(D.F.G.B)	(C.D.D#.G)	(C.D#.G)	28	0.0015704	1	5.7504e-005
45	(D#.G.A#)	(C.D#.F.G#)	(D.F.A#)	28	0.0015704	2	0.00011501
46	(D.F.A#)	(D#.G.A#)	(C.D#.G)	27	0.0015143	13	0.00074756
47	(D#.G.A#)	(D.F.G.A#)	(C.D#.G)	27	0.0015143	1	5.7504e-005
48	(C.F.G#)	(D.F.G#)	(D#.G.A#)	26	0.0014582	1	5.7504e-005
49	(D#.G.A#)	(D.F.A#)	(D#.G.A#)	26	0.0014582	6	0.00034503
50	(D.G.B)	(C.D#.G)	(D.F.A#)	26	0.0014582	17	0.00097757

50 most frequent 3-grams in major using metrical segmentation

				f	p	rf	rp
1	(D.G.B)	(C.E.G)	(C.E.G)	204	0.018311	120	0.010661
2	(C.E.G)	(D.G.B)	(C.E.G)	190	0.017054	116	0.010306
3	(C.E.G)	(C.E.G)	(C.E.G)	153	0.013733	204	0.018124
4	(C.D.F.A)	(D.G.B)	(C.E.G)	141	0.012656	8	0.00071073
5	(C.E.G)	(C.E.G)	(C.F.A)	132	0.011848	59	0.0052416
6	(C.E.G)	(C.E.G)	(D.G.B)	130	0.011669	141	0.012527
7	(D.G.B)	(D.G.B)	(C.E.G)	116	0.010412	73	0.0064854
8	(D.G.B)	(C.E.G)	(C.F.A)	113	0.010143	30	0.0026652
9	(C.E.G)	(C.F.A)	(C.E.G)	106	0.0095144	47	0.0041756
10	(D.G.B)	(C.E.G)	(D.G.B)	91	0.008168	66	0.0058635
11	(C.E.G)	(D.G.B)	(D.G.B)	80	0.0071807	78	0.0069296
12	(C.F.A)	(C.E.G)	(D.G.B)	73	0.0065524	21	0.0018657
13	(C.E.G)	(C.D.F.A)	(D.G.B)	66	0.0059241	7	0.00062189
14	(C.E.G)	(C.F.A)	(D.G.B)	64	0.0057445	38	0.003376
15	(C.E.A)	(D.G.B)	(C.E.G)	62	0.005565	32	0.0028429
16	(D.F#.A)	(D.G.B)	(C.E.G)	60	0.0053855	8	0.00071073
17	(C.D.G)	(D.G.B)	(C.E.G)	60	0.0053855	13	0.0011549
18	(C.E.G)	(D.G.B)	(C.E.A)	58	0.005206	39	0.0034648
19	(C.E.G)	(C.D.G)	(D.G.B)	54	0.004847	8	0.00071073
20	(C.E.G)	(C.E.G)	(C.E.A)	51	0.0045777	59	0.0052416
21	(D.F.G.B)	(C.E.G)	(C.E.G)	51	0.0045777	25	0.002221
22	(C.E.G.A)	(D.F#.A)	(D.G.B)	48	0.0043084	0	0
23	(D.F.G.B)	(C.E.G)	(D.G.B)	42	0.0037699	12	0.0010661
24	(C.E.A)	(C.D.F.A)	(D.G.B)	41	0.0036801	3	0.00026652
25	(C.D.G)	(C.E.G)	(C.E.G)	40	0.0035903	21	0.0018657
26	(D.F#.A)	(D.G.B)	(D.G.B)	40	0.0035903	6	0.00053305
27	(D.G.B)	(C.E.G)	(C.E.A)	40	0.0035903	36	0.0031983
28	(C.E.G)	(C.D.G)	(C.E.G)	38	0.0034108	15	0.0013326
29	(C.F.A)	(D.G.B)	(C.E.G)	37	0.0033211	34	0.0030206
30	(C.E.G)	(D.F.G.B)	(C.E.G)	36	0.0032313	21	0.0018657
31	(D.G.B)	(C.F.A)	(C.E.G)	36	0.0032313	36	0.0031983
32	(C.F.A)	(C.E.G)	(C.E.G)	36	0.0032313	57	0.005064
33	(C.E.G)	(C.F.A)	(C.F.A)	35	0.0031415	22	0.0019545
34	(D.G.B)	(D.F.G.B)	(C.E.G)	33	0.002962	14	0.0012438
35	(D.G.B)	(C.E.G)	(C.D.G)	33	0.002962	12	0.0010661
36	(C.F.A)	(C.E.G)	(C.F.A)	32	0.0028723	23	0.0020434
37	(C.D.G)	(D.F.G.B)	(C.E.G)	32	0.0028723	3	0.00026652
38	(D.G.B)	(C.E.A)	(D.G.B)	31	0.0027825	25	0.002221
39	(C.E.G)	(D.F.A)	(C.E.G)	30	0.0026928	28	0.0024876
40	(C.F.A)	(C.F.A)	(C.E.G)	29	0.002603	19	0.001688
41	(C.E.G)	(C.F.A)	(D.F.B)	28	0.0025132	2	0.00017768
42	(D.F.A)	(C.E.G)	(D.G.B)	28	0.0025132	17	0.0015103
43	(D.G.B)	(D.F#.A)	(D.G.B)	28	0.0025132	6	0.00053305
44	(C.E.G)	(D.G.B)	(C.F.A)	27	0.0024235	37	0.0032871
45	(D.G.B)	(D.G.B)	(D.G.B)	25	0.002244	41	0.0036425
46	(C.E.G)	(C.E.A)	(E.G.B)	25	0.002244	6	0.00053305
47	(C.E.A)	(C.D.F#.A)	(D.G.B)	24	0.0021542	2	0.00017768
48	(C.E.G)	(C.E.G)	(C.D.G)	23	0.0020644	15	0.0013326
49	(C.F.A)	(D.F.G.B)	(C.E.G)	23	0.0020644	7	0.00062189
50	(C.E.G)	(C.F.A)	(D.F.A)	22	0.0019747	8	0.00071073

50 most frequent 3-grams in minor using metrical segmentation

				f	*p*	*rf*	*rp*
1	(D.G.B)	(C.D#.G)	(C.D#.G)	184	0.017696	44	0.0041726
2	(C.D#.G)	(D.G.B)	(C.D#.G)	153	0.014714	40	0.0037933
3	(D.F.A#)	(D#.G.A#)	(D#.G.A#)	95	0.0091364	12	0.001138
4	(C.D#.G)	(C.D#.G)	(C.D#.G)	90	0.0086555	81	0.0076814
5	(C.D.F.G#)	(D.G.B)	(C.D#.G)	89	0.0085593	4	0.00037933
6	(C.D#.G)	(C.D#.G)	(D.G.B)	84	0.0080785	38	0.0036036
7	(D.G.B)	(D.G.B)	(C.D#.G)	78	0.0075014	22	0.0020863
8	(C.D#.G)	(D.F.A#)	(D#.G.A#)	67	0.0064435	11	0.0010431
9	(C.D#.G)	(C.D.F.G#)	(D.G.B)	64	0.006155	5	0.00047416
10	(D#.G.A#)	(D#.G.A#)	(D.F.A#)	63	0.0060589	7	0.00066382
11	(D.F.A#)	(D#.G.A#)	(C.D#.G)	53	0.0050971	21	0.0019915
12	(D#.G.A#)	(D#.G.A#)	(D#.G.A#)	52	0.005001	21	0.0019915
13	(C.D#.F.G#)	(D.F.A#)	(D#.G.A#)	52	0.005001	1	9.4832e-005
14	(D.G.B)	(C.D#.G)	(D.G.B)	48	0.0046163	24	0.002276
15	(C.D#.G)	(C.D.G)	(D.G.B)	47	0.0045201	3	0.0002845
16	(D#.G.A#)	(C.D#.G#)	(D#.G.A#)	47	0.0045201	2	0.00018966
17	(D#.G.A#)	(D.F.A#)	(D#.G.A#)	47	0.0045201	10	0.00094832
18	(D.G.B)	(C.D#.G)	(D.F.A#)	40	0.0038469	13	0.0012328
19	(C.D.F.G#)	(D.G.B)	(C.E.G)	38	0.0036545	2	0.00018966
20	(C.D#.G#)	(D#.G.A#)	(D#.G.A#)	37	0.0035584	9	0.00085349
21	(C.F.A)	(D.F.A#)	(D#.G.A#)	37	0.0035584	1	9.4832e-005
22	(C.D#.G)	(C.D#.G)	(C.D.G)	36	0.0034622	6	0.00056899
23	(C.D#.G)	(D.G.B)	(C.D#.G#)	35	0.003366	12	0.001138
24	(D.G.B)	(C.D#.G)	(C.F.G#)	34	0.0032699	14	0.0013276
25	(C.D.G)	(D.G.B)	(C.D#.G)	33	0.0031737	4	0.00037933
26	(C.D#.G)	(C.D#.G)	(C.D.F.G#)	31	0.0029813	14	0.0013276
27	(D#.G.A#)	(D#.G.A#)	(C.D#.G#)	30	0.0028852	9	0.00085349
28	(C.D#.G)	(C.D#.G)	(C.F.G#)	30	0.0028852	14	0.0013276
29	(D.F.A#)	(D#.G.A#)	(C.D#.G#)	29	0.002789	5	0.00047416
30	(C.D#.G)	(D#.G.A#)	(C.D#.G#)	29	0.002789	9	0.00085349
31	(D.F.A#)	(D#.G.A#)	(D.F.A#)	29	0.002789	6	0.00056899
32	(D.F.A#)	(D.F.A#)	(D#.G.A#)	29	0.002789	9	0.00085349
33	(C.F.G#)	(C.D.F.G#)	(D.G.B)	28	0.0026928	1	9.4832e-005
34	(C.D#.G)	(C.D.G)	(C.D#.G)	28	0.0026928	6	0.00056899
35	(D#.G.A#)	(C.D#.F.G#)	(D.F.A#)	28	0.0026928	1	9.4832e-005
36	(C.D#.G)	(C.D.F.G#)	(C.D#.G)	27	0.0025967	10	0.00094832
37	(C.D#.G)	(C.D#.G)	(D#.G.A#)	27	0.0025967	26	0.0024656
38	(C.F.G#)	(D.G.B)	(C.D#.G)	27	0.0025967	12	0.001138
39	(C.D.G)	(C.D#.G)	(C.D#.G)	26	0.0025005	3	0.0002845
40	(C.D#.G)	(D.G.B)	(D.G.B)	26	0.0025005	31	0.0029398
41	(D.G.B)	(C.D#.G)	(C.F.A)	25	0.0024043	7	0.00066382
42	(D.G.B)	(C.D#.G)	(C.D#.G#)	25	0.0024043	9	0.00085349
43	(C.D#.G)	(C.D#.G)	(D.F.A#)	25	0.0024043	34	0.0032243
44	(C.D#.G)	(C.F.G#)	(D.G.B)	25	0.0024043	12	0.001138
45	(C.D#.G#)	(C.D.F.G#)	(D.G.B)	25	0.0024043	2	0.00018966
46	(C.D#.G)	(C.D#.G)	(D.G.A#)	24	0.0023081	9	0.00085349
47	(D.G.B)	(C.D#.G)	(D.G.A#)	23	0.002212	1	9.4832e-005
48	(C.D#.G)	(C.F.A)	(D.F.A#)	23	0.002212	3	0.0002845
49	(C.D#.G)	(D.F.B)	(C.D#.G)	23	0.002212	1	9.4832e-005
50	(C.D#.G.A#)	(C.F.A)	(D.F.A#)	22	0.0021158	0	0

50 most frequent 3-grams in major using harmonic approximation

				f	*p*	*rf*	*rp*
1	(C.E.G)	(D.G.B)	(C.E.G)	271	0.024325	167	0.014837
2	(D.G.B)	(C.E.G)	(C.E.G)	265	0.023786	176	0.015636
3	(C.E.G)	(C.E.G)	(C.E.G)	160	0.014361	231	0.020522
4	(C.E.G)	(C.E.G)	(C.F.A)	152	0.013643	83	0.0073738
5	(C.E.G)	(C.E.G)	(D.G.B)	147	0.013195	166	0.014748
6	(D.G.B)	(C.E.G)	(D.G.B)	145	0.013015	94	0.0083511
7	(D.G.B)	(D.G.B)	(C.E.G)	138	0.012387	113	0.010039
8	(D.G.B)	(C.E.G)	(C.F.A)	137	0.012297	43	0.0038202
9	(C.E.G)	(C.F.A)	(C.E.G)	135	0.012117	60	0.0053305
10	(C.D.F.A)	(D.G.B)	(C.E.G)	114	0.010232	10	0.00088842
11	(C.E.A)	(D.G.B)	(C.E.G)	109	0.0097837	45	0.0039979
12	(C.F.A)	(C.E.G)	(D.G.B)	99	0.0088861	47	0.0041756
13	(C.E.G)	(D.G.B)	(D.G.B)	97	0.0087066	107	0.009506
14	(C.E.G)	(C.F.A)	(D.G.B)	83	0.00745	46	0.0040867
15	(C.E.G)	(D.F.G.B)	(C.E.G)	83	0.00745	33	0.0029318
16	(C.E.G)	(D.G.B)	(C.E.A)	80	0.0071807	41	0.0036425
17	(C.E.G)	(C.D.F.A)	(D.G.B)	74	0.0066421	7	0.00062189
18	(D.F#.A)	(D.G.B)	(C.E.G)	72	0.0064626	11	0.00097726
19	(C.F.A)	(D.G.B)	(C.E.G)	70	0.0062831	37	0.0032871
20	(D.F.G.B)	(C.E.G)	(C.E.G)	64	0.0057445	34	0.0030206
21	(D.F.G.B)	(C.E.G)	(D.G.B)	62	0.005565	21	0.0018657
22	(D.F.A)	(D.G.B)	(C.E.G)	60	0.0053855	23	0.0020434
23	(C.E.G)	(C.E.G)	(C.E.A)	57	0.0051162	76	0.006752
24	(C.E.G)	(C.D.G)	(D.G.B)	51	0.0045777	4	0.00035537
25	(C.D.G)	(D.G.B)	(C.E.G)	47	0.0042187	9	0.00079957
26	(C.E.G.A)	(D.F#.A)	(D.G.B)	46	0.0041289	2	0.00017768
27	(D.G.B)	(C.E.A)	(D.G.B)	46	0.0041289	30	0.0026652
28	(D.G.B)	(C.E.G)	(C.E.A)	46	0.0041289	42	0.0037313
29	(C.F.A)	(D.F.G.B)	(C.E.G)	44	0.0039494	9	0.00079957
30	(C.F.A)	(C.E.G)	(C.E.G)	44	0.0039494	78	0.0069296
31	(C.F.A)	(C.E.G)	(C.F.A)	43	0.0038596	16	0.0014215
32	(D.F#.A)	(D.G.B)	(D.G.B)	43	0.0038596	9	0.00079957
33	(D.G.B)	(D.F.G.B)	(C.E.G)	42	0.0037699	12	0.0010661
34	(C.E.A)	(C.D.F.A)	(D.G.B)	41	0.0036801	2	0.00017768
35	(C.E.G)	(C.F.A)	(C.F.A)	41	0.0036801	24	0.0021322
36	(C.E.G)	(D.F.A)	(C.E.G)	40	0.0035903	41	0.0036425
37	(D.G.B)	(C.F.A)	(C.E.G)	40	0.0035903	52	0.0046198
38	(D.G.B)	(D.F#.A)	(D.G.B)	38	0.0034108	7	0.00062189
39	(C.E.G)	(C.F.A)	(D.F.A)	35	0.0031415	20	0.0017768
40	(D.F.A)	(C.E.G)	(D.G.B)	35	0.0031415	33	0.0029318
41	(C.E.G)	(D.F.A)	(D.G.B)	34	0.0030518	27	0.0023987
42	(C.E.A)	(C.D.F#.A)	(D.G.B)	34	0.0030518	1	8.8842e-005
43	(C.F.A)	(D.F.B)	(C.E.G)	34	0.0030518	3	0.00026652
44	(C.D.F#.A)	(D.G.B)	(C.E.G)	33	0.002962	13	0.0011549
45	(C.E.G)	(C.F.A)	(D.F.B)	33	0.002962	2	0.00017768
46	(C.F.A)	(C.F.A)	(C.E.G)	33	0.002962	20	0.0017768
47	(D.G.B)	(C.E.G)	(D.F.A)	32	0.0028723	32	0.0028429
48	(C.E.G)	(D.G.B)	(C.F.A)	31	0.0027825	41	0.0036425
49	(C.E.G)	(D.F.B)	(C.E.G)	30	0.0026928	7	0.00062189
50	(E.G.B)	(C.F.A)	(C.E.G)	30	0.0026928	10	0.00088842

50 most frequent 3-grams in minor using harmonic approximation

				f	***p***	***rf***	***rp***
1	(D.G.B)	(C.D#.G)	(C.D#.G)	228	0.021927	56	0.0053106
2	(C.D#.G)	(D.G.B)	(C.D#.G)	223	0.021446	57	0.0054054
3	(D.F.A#)	(D#.G.A#)	(D#.G.A#)	125	0.012022	21	0.0019915
4	(C.D#.G)	(C.D#.G)	(D.G.B)	111	0.010675	68	0.0064486
5	(C.D#.G)	(C.D#.G)	(C.D#.G)	103	0.0099058	101	0.009578
6	(C.D#.G)	(D.F.A#)	(D#.G.A#)	93	0.008944	25	0.0023708
7	(D.G.B)	(D.G.B)	(C.D#.G)	92	0.0088479	38	0.0036036
8	(D#.G.A#)	(D#.G.A#)	(D.F.A#)	82	0.0078861	14	0.0013276
9	(D#.G.A#)	(D.F.A#)	(D#.G.A#)	77	0.0074053	15	0.0014225
10	(D.G.B)	(C.D#.G)	(D.G.B)	72	0.0069244	22	0.0020863
11	(D.F.A#)	(D#.G.A#)	(C.D#.G)	68	0.0065397	17	0.0016121
12	(C.D#.G)	(C.D.F.G#)	(D.G.B)	62	0.0059627	7	0.00066382
13	(D#.G.A#)	(C.D#.G#)	(D#.G.A#)	62	0.0059627	7	0.00066382
14	(C.D#.G)	(C.F.G#)	(D.G.B)	62	0.0059627	24	0.002276
15	(C.D.F.G#)	(D.G.B)	(C.D#.G)	61	0.0058665	6	0.00056899
16	(C.F.G#)	(D.G.B)	(C.D#.G)	61	0.0058665	26	0.0024656
17	(C.F.A)	(D.F.A#)	(D#.G.A#)	60	0.0057703	3	0.0002845
18	(D.G.B)	(C.D#.G)	(C.F.G#)	58	0.005578	26	0.0024656
19	(D.G.B)	(C.D#.G)	(D.F.A#)	56	0.0053857	27	0.0025605
20	(D#.G.A#)	(D#.G.A#)	(D#.G.A#)	54	0.0051933	22	0.0020863
21	(D.F.A#)	(D#.G.A#)	(D.F.A#)	47	0.0045201	9	0.00085349
22	(C.D#.G)	(D.G.B)	(C.D#.G#)	46	0.0044239	6	0.00056899
23	(C.D#.F.G#)	(D.F.A#)	(D#.G.A#)	45	0.0043278	1	9.4832e-005
24	(D.F.A#)	(D#.G.A#)	(C.D#.G#)	44	0.0042316	5	0.00047416
25	(C.D#.G)	(C.D#.G)	(C.F.G#)	42	0.0040392	42	0.0039829
26	(C.D#.G)	(C.D.G)	(D.G.B)	41	0.0039431	4	0.00037933
27	(C.F.G#)	(C.D#.G)	(D.G.B)	41	0.0039431	21	0.0019915
28	(D#.G.A#)	(D#.G.A#)	(C.D#.G#)	41	0.0039431	6	0.00056899
29	(C.D#.G#)	(D#.G.A#)	(D#.G.A#)	40	0.0038469	6	0.00056899
30	(C.D#.G)	(D.G.B)	(D.G.B)	39	0.0037507	32	0.0030346
31	(D.F.A#)	(D.F.A#)	(D#.G.A#)	38	0.0036545	12	0.001138
32	(C.D#.G)	(D.F.G.B)	(C.D#.G)	37	0.0035584	11	0.0010431
33	(D#.G.A#)	(D.F.A#)	(C.D#.G)	34	0.0032699	22	0.0020863
34	(C.D#.G)	(C.F.A)	(D.F.A#)	33	0.0031737	10	0.00094832
35	(C.D#.G)	(D.F.B)	(C.D#.G)	33	0.0031737	3	0.0002845
36	(D.G.B)	(C.D#.G)	(D.G.A#)	32	0.0030775	13	0.0012328
37	(C.D#.G)	(D#.G.A#)	(C.D#.G#)	32	0.0030775	15	0.0014225
38	(C.D#.G)	(D.G.B)	(C.E.G)	31	0.0029813	12	0.001138
39	(D.F.G.B)	(C.D#.G)	(C.D#.G)	31	0.0029813	14	0.0013276
40	(D.F.G.B)	(C.D#.G)	(D.G.B)	31	0.0029813	6	0.00056899
41	(C.D#.G)	(C.D#.G)	(D.G.A#)	30	0.0028852	26	0.0024656
42	(C.D#.G)	(C.D#.G)	(D#.G.A#)	30	0.0028852	71	0.006733
43	(D#.G.A#)	(C.F.A)	(D.F.A#)	30	0.0028852	3	0.0002845
44	(C.D#.G)	(C.D#.G)	(D.F.A#)	30	0.0028852	43	0.0040778
45	(C.F.G#)	(D.G.B)	(D.G.B)	29	0.002789	14	0.0013276
46	(D.G.B)	(C.D#.G)	(C.F.A)	29	0.002789	12	0.001138
47	(D#.G.A#)	(C.D#.F.G#)	(D.F.A#)	29	0.002789	1	9.4832e-005
48	(C.D.F.G#)	(D.G.B)	(C.E.G)	29	0.002789	1	9.4832e-005
49	(D.G.B)	(C.D#.G)	(C.D#.G#)	28	0.0026928	9	0.00085349
50	(C.D.G)	(D.G.B)	(C.D#.G)	26	0.0025005	4	0.00037933

50 most frequent 3-grams in major within a random corpus

				f	*p*
1	(C.E.G)	(C.E.G)	(C.E.G)	107	0.0056632
2	(D.G.B)	(C.E.G)	(C.E.G)	88	0.0046576
3	(C.E.G)	(D.G.B)	(C.E.G)	88	0.0046576
4	(C.E.G)	(C.E.G)	(D.G.B)	83	0.0043929
5	(D.G.B)	(D.G.B)	(C.E.G)	53	0.0028051
6	(D.G.B)	(C.E.G)	(D.G.B)	52	0.0027522
7	(C.E.G)	(D.G.B)	(D.G.B)	50	0.0026463
8	(C.F.A)	(C.E.G)	(C.E.G)	46	0.0024346
9	(C.E.G)	(C.E.G)	(C.E.A)	43	0.0022759
10	(C.E.G)	(C.F.A)	(C.E.G)	41	0.00217
11	(C.E.G)	(C.E.G)	(C.F.A)	38	0.0020112
12	(D.G.B)	(D.G.B)	(D.G.B)	36	0.0019054
13	(C.E.G)	(C.F.A)	(D.G.B)	34	0.0017995
14	(C.E.G)	(D.F.G.B)	(C.E.G)	33	0.0017466
15	(C.F.A)	(C.E.G)	(D.G.B)	32	0.0016937
16	(D.F.G.B)	(C.E.G)	(C.E.G)	30	0.0015878
17	(C.E.G)	(C.E.A)	(C.E.G)	29	0.0015349
18	(D.G.B)	(C.F.A)	(C.E.G)	27	0.001429
19	(D.G.B)	(C.E.A)	(C.E.G)	26	0.0013761
20	(C.E.G)	(D.G.B)	(C.E.A)	26	0.0013761
21	(C.E.G)	(D.G.B)	(C.F.A)	25	0.0013232
22	(C.F.A)	(D.G.B)	(C.E.G)	25	0.0013232
23	(D.G.B)	(C.E.G)	(D.F.G.B)	24	0.0012702
24	(C.E.A)	(C.E.G)	(C.E.G)	24	0.0012702
25	(D.G.B)	(C.E.G)	(C.F.A)	23	0.0012173
26	(D.G.B)	(C.E.G)	(C.E.A)	23	0.0012173
27	(C.E.G)	(C.E.G)	(D.F.A)	22	0.0011644
28	(C.E.G)	(D.G.B)	(D.F.G.B)	22	0.0011644
29	(C.E.G)	(C.E.A)	(D.G.B)	22	0.0011644
30	(D.F.G.B)	(D.G.B)	(C.E.G)	21	0.0011115
31	(E.G.B)	(C.E.G)	(C.E.G)	21	0.0011115
32	(C.E.A)	(C.E.G)	(D.G.B)	21	0.0011115
33	(D.F.G.B)	(C.E.G)	(D.G.B)	20	0.0010585
34	(D.G.B)	(D.F.G.B)	(C.E.G)	20	0.0010585
35	(C.E.G)	(C.E.G)	(D.F.G.B)	20	0.0010585
36	(D.F.A)	(D.G.B)	(C.E.G)	19	0.0010056
37	(C.E.G)	(D.F.G.B)	(D.G.B)	19	0.0010056
38	(C.E.A)	(D.G.B)	(C.E.G)	19	0.0010056
39	(C.E.G)	(D.F.A)	(D.G.B)	18	0.00095268
40	(D.F.A)	(C.E.G)	(C.E.G)	18	0.00095268
41	(D.G.B)	(E.G.B)	(C.E.G)	17	0.00089976
42	(C.E.G)	(C.E.G)	(E.G.B)	17	0.00089976
43	(C.E.A)	(D.G.B)	(D.G.B)	17	0.00089976
44	(C.E.G)	(C.E.G)	(C.D.F.A)	17	0.00089976
45	(C.E.G)	(E.G.B)	(C.E.G)	17	0.00089976
46	(C.D.F#.A)	(C.E.G)	(C.E.G)	17	0.00089976
47	(C.F.A)	(C.E.G)	(C.E.A)	17	0.00089976
48	(C.E.G)	(C.E.G)	(D.F.B)	16	0.00084683
49	(C.E.G)	(C.F.A)	(C.F.A)	16	0.00084683
50	(D.F#.A)	(C.E.G)	(C.E.G)	16	0.00084683

50 most frequent 3-grams in minor within a random corpus

				f	***p***
1	(C.D#.G)	(C.D#.G)	(C.D#.G)	55	0.0031627
2	(D#.G.A#)	(C.D#.G)	(C.D#.G)	37	0.0021277
3	(C.D#.G)	(D#.G.A#)	(D#.G.A#)	26	0.0014951
4	(C.D#.G)	(D.F.A#)	(C.D#.G)	26	0.0014951
5	(C.D#.G)	(D.G.B)	(C.D#.G)	25	0.0014376
6	(C.D#.G)	(C.D#.G)	(D.G.B)	25	0.0014376
7	(D.F.A#)	(C.D#.G)	(C.D#.G)	25	0.0014376
8	(D.G.B)	(D#.G.A#)	(C.D#.G)	22	0.0012651
9	(D.G.B)	(C.D#.G)	(C.D#.G)	22	0.0012651
10	(C.D#.G)	(C.D#.G)	(D#.G.A#)	21	0.0012076
11	(D#.G.A#)	(C.D#.G)	(D.G.B)	20	0.0011501
12	(D.G.B)	(D.F.A#)	(C.D#.G)	20	0.0011501
13	(C.D#.G)	(D.G.B)	(D.F.A#)	20	0.0011501
14	(C.D#.G)	(D#.G.A#)	(C.D#.G)	19	0.0010926
15	(D#.G.A#)	(D#.G.A#)	(C.D#.G)	18	0.0010351
16	(C.D#.G)	(C.D#.G)	(C.F.G#)	18	0.0010351
17	(D.G.B)	(C.D#.G)	(D.F.A#)	17	0.00097757
18	(C.D#.G)	(C.D#.G)	(C.D.F.G#)	17	0.00097757
19	(D.F.A#)	(D.F.A#)	(C.D#.G)	17	0.00097757
20	(C.D#.G)	(D.G.B)	(D#.G.A#)	16	0.00092007
21	(C.F.G#)	(C.D#.G)	(D#.G.A#)	16	0.00092007
22	(C.F.G#)	(C.D#.G)	(C.D#.G)	16	0.00092007
23	(D.G.B)	(C.D#.G)	(D#.G.A#)	16	0.00092007
24	(C.D#.G)	(D#.G.A#)	(D.G.B)	16	0.00092007
25	(D.G.B)	(D.G.B)	(C.D#.G)	15	0.00086256
26	(D#.G.A#)	(C.D#.G)	(D#.G.A#)	14	0.00080506
27	(C.D#.G)	(C.D#.G)	(D.F.A#)	14	0.00080506
28	(D#.G.A#)	(D.G.B)	(C.D#.G)	14	0.00080506
29	(D.F.A#)	(C.D#.G)	(D#.G.A#)	14	0.00080506
30	(D.F.A#)	(C.D#.G)	(D.G.B)	14	0.00080506
31	(C.D#.G)	(C.F.G#)	(C.D#.G)	13	0.00074756
32	(C.D#.G)	(D#.G.A#)	(D.F.A#)	13	0.00074756
33	(C.D#.G)	(D.F.A#)	(D.G.B)	13	0.00074756
34	(C.D#.G)	(C.D#.G)	(C.F.A)	13	0.00074756
35	(D.F.A#)	(D#.G.A#)	(C.D#.G)	13	0.00074756
36	(D#.G.A#)	(D.G.B)	(D#.G.A#)	12	0.00069005
37	(C.D#.G)	(D.G.A#)	(C.D#.G)	12	0.00069005
38	(D#.G.A#)	(C.D#.G)	(D.F.A#)	12	0.00069005
39	(C.D#.G)	(C.D#.G)	(C.D#.G#)	12	0.00069005
40	(C.D#.G)	(C.D#.G#)	(C.D#.G)	12	0.00069005
41	(D.G.B)	(C.D#.G)	(D.F.G.B)	12	0.00069005
42	(C.D#.G)	(D.F.A#)	(C.F.G#)	12	0.00069005
43	(C.D#.G)	(D#.G.A#)	(C.F.G#)	12	0.00069005
44	(C.D#.G)	(C.D.G)	(C.D#.G)	12	0.00069005
45	(C.D#.G)	(D.F.A#)	(D.F.A#)	11	0.00063255
46	(D#.G.A#)	(D#.G.A#)	(D.G.B)	11	0.00063255
47	(C.F.G#)	(D#.G.A#)	(C.D#.G)	11	0.00063255
48	(C.D#.G)	(D.F.G.B)	(D.G.B)	11	0.00063255
49	(C.D#.G)	(D.G.B)	(D.G.B)	11	0.00063255
50	(D.G.B)	(D.F.G.B)	(C.D#.G)	11	0.00063255

50 most frequent 4-grams in major using dense segmentation

					f	***p***	***rf***	***rp***
1	(D.G.B)	(D.F.G.B)	(C.E.G)	(C.E.G)	80	0.0044265	4	0.00021398
2	(D.G.B)	(D.G.B)	(D.F.G.B)	(C.E.G)	44	0.0024346	2	0.00010699
3	(C.D.F.A)	(D.G.B)	(D.F.G.B)	(C.E.G)	43	0.0023792	0	0
4	(C.E.G)	(C.F.A)	(C.E.F.A)	(D.G.B)	42	0.0023239	1	5.3496e-005
5	(C.D.G)	(D.G.B)	(D.F.G.B)	(C.E.G)	38	0.0021026	0	0
6	(C.E.G)	(C.D.F.A)	(D.G.B)	(C.E.G)	32	0.0017706	0	0
7	(C.E.G)	(D.G.B)	(D.F.G.B)	(C.E.G)	32	0.0017706	3	0.00016049
8	(C.E.G)	(C.E.G)	(D.G.B)	(C.E.G)	30	0.0016599	17	0.00090943
9	(C.E.G)	(C.D.F.A)	(D.G.B)	(D.F.G.B)	29	0.0016046	0	0
10	(D.G.B)	(C.E.G)	(C.E.G.B)	(C.F.A)	28	0.0015493	3	0.00016049
11	(D.G.B)	(D.F.G.B)	(C.E.G)	(D.G.B)	28	0.0015493	4	0.00021398
12	(C.D.G)	(D.G.B)	(C.E.G)	(C.E.G)	27	0.0014939	1	5.3496e-005
13	(D.F.G.B)	(C.E.G)	(C.E.G)	(D.G.B)	27	0.0014939	3	0.00016049
14	(C.E.G)	(C.E.G)	(C.D.F.G)	(C.E.G)	26	0.0014386	2	0.00010699
15	(C.D.E.G)	(C.E.G)	(C.D.G)	(D.G.B)	25	0.0013833	0	0
16	(C.E.G)	(C.E.F.G)	(C.D.G)	(D.G.B)	24	0.0013279	1	5.3496e-005
17	(C.E.G.A)	(D.F#.A)	(C.D.F#.A)	(D.G.B)	24	0.0013279	0	0
18	(C.E.G)	(C.E.G)	(C.F.A)	(C.E.F.A)	24	0.0013279	0	0
19	(C.D.F.A)	(D.G.B)	(C.E.G)	(C.E.G)	22	0.0012173	4	0.00021398
20	(D.F.B)	(C.E.G)	(C.E.F.G)	(D.G.B)	21	0.001162	0	0
21	(C.E.G)	(D.F.A)	(D.F.B)	(C.E.G)	21	0.001162	0	0
22	(C.E.G)	(C.D.G)	(D.G.B)	(C.E.G)	21	0.001162	0	0
23	(D.F#.A)	(C.D.F#.A)	(D.G.B)	(C.E.G)	21	0.001162	1	5.3496e-005
24	(C.E.G)	(C.E.F.G)	(D.G.B)	(D.F.G.B)	20	0.0011066	0	0
25	(D.G.B)	(D.F.G.B)	(C.E.G)	(C.D.E.G)	20	0.0011066	0	0
26	(D.G.B)	(C.F.A)	(D.F.A.B)	(C.E.G)	20	0.0011066	1	5.3496e-005
27	(D.G.B)	(D.F.G.B)	(C.E.G)	(C.F.A)	20	0.0011066	1	5.3496e-005
28	(C.E.G)	(C.E.G)	(C.E.G.B)	(C.E.A)	20	0.0011066	0	0
29	(C.E.G)	(C.D.F)	(D.F.B)	(C.E.G)	19	0.0010513	0	0
30	(C.E.G)	(C.D.G)	(D.G.B)	(D.F.G.B)	19	0.0010513	0	0
31	(C.E.G)	(D.G.B)	(C.E.G)	(D.G.B)	19	0.0010513	14	0.00074894
32	(C.E.G)	(D.G.B)	(C.E.G)	(C.E.G.B)	19	0.0010513	1	5.3496e-005
33	(C.F.A)	(C.E.F.A)	(D.G.B)	(C.F.A)	19	0.0010513	1	5.3496e-005
34	(D.F#.A)	(C.D.F#.A)	(D.G.B)	(D.G.B)	19	0.0010513	0	0
35	(C.E.G)	(C.D.E.G)	(C.E.G)	(C.E.F.G)	18	0.00099596	0	0
36	(D.F.A)	(D.F.B)	(C.E.G)	(C.E.F.G)	18	0.00099596	0	0
37	(D.G.B)	(C.E.G)	(C.D.F.G)	(C.E.G)	18	0.00099596	0	0
38	(C.E.G)	(C.E.G)	(C.D.E.G)	(C.E.G)	18	0.00099596	1	5.3496e-005
39	(C.D.F.A)	(D.F.B)	(D.F.A)	(D.G.B)	17	0.00094063	0	0
40	(D.G.B)	(C.G.A)	(C.F#.A)	(D.G.B)	17	0.00094063	0	0
41	(D.G.B)	(C.E.G.A)	(D.F#.A)	(C.D.F#.A)	17	0.00094063	0	0
42	(C.E.G)	(C.E.G.B)	(C.E.A)	(C.E.G.A)	17	0.00094063	0	0
43	(C.E.G)	(C.D.F.A)	(D.F.B)	(D.F.A)	16	0.0008853	0	0
44	(D.G.B)	(C.E.G)	(C.E.G)	(C.E.G.B)	16	0.0008853	0	0
45	(D.F.G.B)	(C.E.G)	(C.E.G)	(C.F.A)	16	0.0008853	2	0.00010699
46	(C.D.F)	(D.F.B)	(C.E.G)	(C.E.F.G)	16	0.0008853	0	0
47	(C.E.G)	(C.E.G.B)	(C.F.A)	(C.E.G)	16	0.0008853	2	0.00010699
48	(C.E.F.A)	(D.G.B)	(D.F.G.B)	(C.E.G)	16	0.0008853	0	0
49	(C.F.A)	(D.F.A.B)	(C.E.G)	(C.E.F.A)	16	0.0008853	0	0
50	(C.E.G)	(C.E.F.G)	(D.G.B)	(D.G.B)	15	0.00082997	0	0

50 most frequent 4-grams in minor using dense segmentation

					f	***p***	***rf***	***rp***
1	(D.G.B)	(D.F.G.B)	(C.D#.G)	(C.D#.G)	96	0.0054406	0	0
2	(C.D#.G)	(C.D.F.G#)	(D.G.B)	(D.F.G.B)	52	0.002947	0	0
3	(C.D#.G)	(D.G.B)	(D.F.G.B)	(C.D#.G)	45	0.0025503	1	5.8123e-005
4	(C.D.F.G#)	(D.G.B)	(D.F.G.B)	(C.D#.G)	42	0.0023803	0	0
5	(D.F.A#)	(D.F.G#.A#)	(D#.G.A#)	(D#.G.A#)	37	0.0020969	0	0
6	(D.F.G.B)	(C.D#.G)	(C.D#.G)	(C.D#.G.A#)	22	0.0012468	0	0
7	(D.G.B)	(D.G.B)	(D.F.G.B)	(C.D#.G)	21	0.0011901	0	0
8	(C.F.A)	(C.D#.F.A)	(D.F.A#)	(D#.G.A#)	21	0.0011901	0	0
9	(C.D#.G)	(D.G.B)	(C.D#.G)	(D.G.B)	21	0.0011901	1	5.8123e-005
10	(C.D.G)	(D.G.B)	(D.F.G.B)	(C.D#.G)	20	0.0011335	1	5.8123e-005
11	(C.D#.G)	(C.D.D#.G)	(C.D#.G)	(C.D#.F.G)	20	0.0011335	1	5.8123e-005
12	(C.D#.G)	(C.D#.F.G)	(D.G.B)	(D.F.G.B)	20	0.0011335	0	0
13	(C.D.F.G#)	(D.G.B)	(D.F.G.B)	(C.E.G)	20	0.0011335	0	0
14	(C.F.G#)	(D.G.B)	(D.G.B)	(C.D#.G)	19	0.0010768	0	0
15	(C.D#.G)	(C.D.F)	(D.F.B)	(C.D#.G)	19	0.0010768	0	0
16	(C.D#.G)	(C.D#.G)	(C.D.F.G)	(C.D#.G)	19	0.0010768	1	5.8123e-005
17	(C.D#.G)	(C.D#.F.G)	(C.D.G)	(D.G.B)	18	0.0010201	0	0
18	(D.F.A#)	(D#.G.A#)	(D#.G.A#)	(D.F.A#)	18	0.0010201	2	0.00011625
19	(C.D#.F.G#)	(D.F.A#)	(D.F.G#.A#)	(D#.G.A#)	18	0.0010201	0	0
20	(C.D#.G)	(C.D#.G)	(D.G.B)	(C.D#.G)	18	0.0010201	4	0.00023249
21	(C.F.G#)	(D.G.B)	(D.F.G.B)	(C.D#.G)	17	0.00096345	1	5.8123e-005
22	(D.G.B)	(C.D#.G)	(C.D#.G.A#)	(C.F.G#)	17	0.00096345	0	0
23	(C.D.G)	(D.G.B)	(C.D#.G)	(C.D#.G)	16	0.00090677	0	0
24	(C.D#.G)	(D.G.B)	(D.F.G.B)	(C.D#.G#)	16	0.00090677	0	0
25	(D.F.A#)	(D.F.A#)	(D.F.G#.A#)	(D#.G.A#)	16	0.00090677	0	0
26	(D#.F.A#)	(D.F.A#)	(D.F.G#.A#)	(D#.G.A#)	15	0.0008501	0	0
27	(C.D#.G)	(C.D.F.G#)	(C.D.F)	(D.G.B)	15	0.0008501	0	0
28	(C.D#.G)	(C.D#.G)	(C.D.G)	(D.G.B)	15	0.0008501	0	0
29	(C.D#.G)	(C.D#.A)	(D.F.A#)	(D.F.G#.A#)	15	0.0008501	0	0
30	(C.D#.G)	(C.D#.G)	(C.D#.G.A#)	(C.D#.G#)	15	0.0008501	0	0
31	(D.G.B)	(D.F.G.B)	(C.D#.G)	(D.G.B)	15	0.0008501	0	0
32	(C.D.F.G#)	(C.D.F)	(D.G.B)	(D.G.B)	14	0.00079343	0	0
33	(C.D.G)	(D.G.B)	(D.G.A)	(D.G.B)	14	0.00079343	0	0
34	(C.D.F.G)	(C.D#.G)	(D.G.B)	(C.D#.G)	14	0.00079343	1	5.8123e-005
35	(C.D#.G)	(C.D.F.G)	(C.D#.G)	(D.G.B)	14	0.00079343	1	5.8123e-005
36	(D.G.B)	(C.D#.G)	(D.G.B)	(C.D#.G)	14	0.00079343	1	5.8123e-005
37	(C.D#.G)	(C.D#.G)	(C.D.D#.G)	(C.D#.G)	14	0.00079343	2	0.00011625
38	(C.D#.G)	(C.D.G)	(D.G.B)	(D.F.G.B)	14	0.00079343	0	0
39	(C.D#.G)	(C.D.D#.G)	(C.D.G)	(D.G.B)	13	0.00073675	0	0
40	(C.D.F.G#)	(C.F.G#)	(D.G.B)	(D.G.B)	13	0.00073675	0	0
41	(D.G.B)	(C.D#.G)	(C.D.D#.G)	(C.D#.G)	13	0.00073675	0	0
42	(D.G.B)	(C.D#.G)	(C.D#.G.A#)	(C.F.A)	13	0.00073675	0	0
43	(D.G.B)	(C.D#.G)	(D.G.B)	(D.F.G.B)	13	0.00073675	0	0
44	(C.D#.G)	(D.G.B)	(D.G.B)	(C.D#.G)	13	0.00073675	0	0
45	(C.D#.G)	(C.D.F.G#)	(C.F.G#)	(D.G.B)	13	0.00073675	0	0
46	(D.G.B)	(C.D#.G)	(D.F.A#)	(D#.G.A#)	13	0.00073675	1	5.8123e-005
47	(D.F.G.B)	(C.D#.G)	(D.G.B)	(C.D#.G)	13	0.00073675	0	0
48	(D#.G.A#)	(D.D#.G.A#)	(C.D#.G)	(C.D#.G.A#)	13	0.00073675	0	0
49	(D.F.A#)	(D#.G.A#)	(D.D#.G.A#)	(C.D#.G)	13	0.00073675	0	0
50	(C.D#.G)	(C.D.G)	(D.G.B)	(C.D#.G)	12	0.00068008	0	0

50 most frequent 4-grams in major using metrical segmentation

					f	*p*	*rf*	*rp*
1	(C.E.G)	(C.D.F.A)	(D.G.B)	(C.E.G)	58	0.0053016	4	0.00036183
2	(C.D.F.A)	(D.G.B)	(C.E.G)	(C.E.G)	55	0.0050274	1	9.0457e-005
3	(D.G.B)	(C.E.G)	(C.E.G)	(C.E.G)	52	0.0047532	28	0.0025328
4	(C.E.G)	(C.E.G)	(D.G.B)	(C.E.G)	49	0.004479	29	0.0026232
5	(D.G.B)	(C.E.G)	(C.F.A)	(C.E.G)	45	0.0041133	8	0.00072365
6	(D.G.B)	(C.E.G)	(C.E.G)	(C.F.A)	43	0.0039305	5	0.00045228
7	(C.E.G)	(D.G.B)	(C.E.G)	(C.E.G)	42	0.0038391	33	0.0029851
8	(C.E.A)	(C.D.F.A)	(D.G.B)	(C.E.G)	37	0.0033821	0	0
9	(C.E.G)	(D.G.B)	(D.G.B)	(C.E.G)	36	0.0032907	24	0.002171
10	(C.E.G)	(C.D.G)	(D.G.B)	(C.E.G)	34	0.0031079	3	0.00027137
11	(C.E.G)	(C.E.G)	(C.E.G)	(C.E.G)	32	0.002925	47	0.0042515
12	(C.E.G)	(C.F.A)	(C.E.G)	(D.G.B)	32	0.002925	2	0.00018091
13	(C.F.A)	(C.E.G)	(D.G.B)	(C.E.G)	31	0.0028336	5	0.00045228
14	(C.E.G)	(D.G.B)	(C.E.G)	(C.F.A)	31	0.0028336	7	0.0006332
15	(C.E.G)	(C.E.G)	(C.E.G)	(C.F.A)	30	0.0027422	12	0.0010855
16	(D.G.B)	(C.E.G)	(C.E.G)	(C.E.A)	30	0.0027422	10	0.00090457
17	(C.E.G)	(C.E.G)	(C.E.G)	(D.G.B)	30	0.0027422	33	0.0029851
18	(D.G.B)	(C.E.G)	(C.E.G)	(D.G.B)	29	0.0026508	23	0.0020805
19	(C.E.G)	(C.E.G)	(C.F.A)	(D.G.B)	27	0.002468	5	0.00045228
20	(D.G.B)	(D.G.B)	(C.E.G)	(C.E.G)	27	0.002468	16	0.0014473
21	(D.F#.A)	(D.G.B)	(D.G.B)	(C.E.G)	24	0.0021938	1	9.0457e-005
22	(D.G.B)	(C.E.G)	(C.F.A)	(D.G.B)	23	0.0021024	5	0.00045228
23	(C.E.G)	(D.G.B)	(C.E.G)	(D.G.B)	22	0.002011	16	0.0014473
24	(C.E.G)	(C.E.G)	(C.F.A)	(C.E.G)	22	0.002011	13	0.0011759
25	(C.E.G.A)	(D.F#.A)	(D.G.B)	(C.E.G)	21	0.0019196	0	0
26	(D.G.B)	(C.E.G)	(D.G.B)	(C.E.G)	20	0.0018282	22	0.00199
27	(C.F.A)	(D.G.B)	(C.F.A)	(C.E.G)	20	0.0018282	5	0.00045228
28	(C.D.E.G)	(C.D.F.A)	(D.G.B)	(C.E.G)	19	0.0017367	0	0
29	(C.E.G)	(C.F.A)	(D.G.B)	(C.F.A)	19	0.0017367	3	0.00027137
30	(D.G.B)	(C.E.G)	(D.G.B)	(D.G.B)	19	0.0017367	12	0.0010855
31	(D.G.B)	(C.E.G.A)	(D.F#.A)	(D.G.B)	18	0.0016453	0	0
32	(C.E.G)	(C.F.A)	(D.F.B)	(C.E.G)	18	0.0016453	2	0.00018091
33	(C.E.G)	(C.D.G)	(D.F.G.B)	(C.E.G)	18	0.0016453	0	0
34	(D.F.G.B)	(C.E.G)	(C.E.G)	(C.E.G)	17	0.0015539	6	0.00054274
35	(C.E.G)	(C.E.G)	(C.F.A)	(D.F.A)	16	0.0014625	3	0.00027137
36	(C.E.A)	(D.G.B)	(C.E.G)	(C.F.A)	16	0.0014625	2	0.00018091
37	(C.E.G)	(C.E.G)	(D.G.B)	(C.E.A)	16	0.0014625	12	0.0010855
38	(C.E.G)	(D.G.B)	(C.E.G)	(C.D.G)	16	0.0014625	3	0.00027137
39	(C.E.G)	(D.G.B)	(C.E.A)	(D.G.B)	16	0.0014625	7	0.0006332
40	(C.E.G)	(C.E.G)	(D.G.B)	(D.G.B)	16	0.0014625	32	0.0028946
41	(D.F.G.B)	(C.E.G)	(D.G.B)	(C.E.G)	16	0.0014625	1	9.0457e-005
42	(C.D.G)	(D.F.G.B)	(C.E.G)	(C.E.G)	16	0.0014625	0	0
43	(C.E.G)	(C.F.A)	(C.E.G)	(C.F.A)	15	0.0013711	4	0.00036183
44	(C.F.A)	(C.D.F.A)	(D.G.B)	(C.E.G)	15	0.0013711	1	9.0457e-005
45	(E.G#.B)	(C.E.A)	(D.G.B)	(C.E.G)	15	0.0013711	0	0
46	(D.G.B)	(C.E.G)	(C.D.G)	(D.G.B)	15	0.0013711	1	9.0457e-005
47	(C.D.F#.A)	(D.G.B)	(D.G.B)	(C.E.G)	14	0.0012797	1	9.0457e-005
48	(D.F#.A)	(D.G.B)	(C.E.G)	(C.F.A)	14	0.0012797	0	0
49	(C.E.G)	(C.F.A)	(C.F.A)	(C.E.G)	14	0.0012797	5	0.00045228
50	(D.G.B)	(D.G.B)	(C.E.G)	(C.F.A)	13	0.0011883	6	0.00054274

50 most frequent 4-grams in minor using metrical segmentation

					f	***p***	***rf***	***rp***
1	(C.D.F.G#)	(D.G.B)	(C.D#.G)	(C.D#.G)	52	0.0050915	1	9.6525e-005
2	(C.D#.G)	(C.D#.G)	(D.G.B)	(C.D#.G)	47	0.004602	11	0.0010618
3	(C.D#.G)	(D.G.B)	(C.D#.G)	(C.D#.G)	42	0.0041124	8	0.0007722
4	(C.D#.G)	(C.D.F.G#)	(D.G.B)	(C.D#.G)	40	0.0039166	0	0
5	(D.G.B)	(C.D#.G)	(C.D#.G)	(C.D#.G)	35	0.003427	8	0.0007722
6	(D.G.B)	(C.D#.G)	(D.F.A#)	(D#.G.A#)	29	0.0028395	3	0.00028958
7	(D.G.B)	(C.D#.G)	(D.G.B)	(C.D#.G)	27	0.0026437	6	0.00057915
8	(D#.G.A#)	(C.D#.G#)	(D#.G.A#)	(D#.G.A#)	26	0.0025458	0	0
9	(D.F.A#)	(D#.G.A#)	(D#.G.A#)	(D.F.A#)	26	0.0025458	0	0
10	(D#.G.A#)	(D#.G.A#)	(D.F.A#)	(D#.G.A#)	24	0.0023499	0	0
11	(C.D.G)	(D.G.B)	(C.D#.G)	(C.D#.G)	23	0.002252	0	0
12	(C.D#.G)	(C.D.G)	(D.G.B)	(C.D#.G)	23	0.002252	0	0
13	(C.D#.F.G#)	(D.F.A#)	(D#.G.A#)	(D#.G.A#)	23	0.002252	0	0
14	(D#.G.A#)	(C.D#.F.G#)	(D.F.A#)	(D#.G.A#)	22	0.0021541	0	0
15	(C.D#.G)	(D.G.B)	(C.D#.G)	(D.G.B)	22	0.0021541	5	0.00048263
16	(C.D#.G)	(D#.G.A#)	(C.D#.G#)	(D#.G.A#)	21	0.0020562	0	0
17	(C.D#.G)	(C.F.G#)	(C.D.F.G#)	(D.G.B)	20	0.0019583	0	0
18	(D#.G.A#)	(D.F.A#)	(D#.G.A#)	(D#.G.A#)	19	0.0018604	1	9.6525e-005
19	(C.D#.G)	(C.D#.G)	(C.D#.G)	(D.G.B)	19	0.0018604	7	0.00067568
20	(C.D#.G)	(D.G.B)	(D.G.B)	(C.D#.G)	18	0.0017625	10	0.00096525
21	(C.D#.G#)	(D#.G.A#)	(D#.G.A#)	(D#.G.A#)	17	0.0016645	1	9.6525e-005
22	(C.D#.G)	(C.D#.F.G#)	(D.F.A#)	(D#.G.A#)	17	0.0016645	0	0
23	(D.G.B)	(D.G.B)	(C.D#.G)	(C.D#.G)	16	0.0015666	6	0.00057915
24	(D#.G.A#)	(D#.G.A#)	(D#.G.A#)	(D.F.A#)	16	0.0015666	3	0.00028958
25	(C.D#.G#)	(C.D.F.G#)	(D.G.B)	(C.D#.G)	16	0.0015666	0	0
26	(D.G.B)	(C.D#.G)	(C.F.A)	(D.F.A#)	16	0.0015666	0	0
27	(D.G.B)	(C.D#.G)	(C.D#.G)	(D.G.B)	16	0.0015666	3	0.00028958
28	(C.D#.G)	(C.F.G#)	(D.G.B)	(C.D#.G)	15	0.0014687	2	0.00019305
29	(C.F.G#)	(C.D.F.G#)	(D.G.B)	(C.E.G)	15	0.0014687	0	0
30	(D.G.B)	(C.D#.G)	(C.D#.G)	(D#.G.A#)	15	0.0014687	4	0.0003861
31	(D.F.A#)	(D#.G.A#)	(D#.G.A#)	(D#.G.A#)	15	0.0014687	2	0.00019305
32	(C.D#.G)	(C.D#.G)	(D.F.A#)	(D#.G.A#)	15	0.0014687	2	0.00019305
33	(D.G.B)	(D.G.B)	(C.D#.G)	(D.F.A#)	14	0.0013708	0	0
34	(C.D#.G)	(D.G.B)	(C.D#.G)	(C.F.G#)	14	0.0013708	3	0.00028958
35	(C.D#.G)	(C.D#.G)	(C.D#.G)	(C.D#.G)	13	0.0012729	14	0.0013514
36	(D.F.G#)	(D.G.B)	(D.G.B)	(D.G.A#)	13	0.0012729	0	0
37	(D.G.B)	(C.D#.G)	(C.D.F.G#)	(D.G.B)	13	0.0012729	1	9.6525e-005
38	(D.G.B)	(C.D#.G)	(C.D#.G)	(C.D.G)	13	0.0012729	0	0
39	(C.D#.F.G#)	(D.F.A#)	(D#.G.A#)	(C.D#.G)	13	0.0012729	0	0
40	(D.G.B)	(C.D#.G)	(C.D#.G)	(C.D#.G#)	13	0.0012729	3	0.00028958
41	(D.G.B)	(C.D#.G)	(C.D#.G)	(C.F.A)	13	0.0012729	1	9.6525e-005
42	(C.D#.G)	(D.F.A#)	(D#.G.A#)	(C.D#.G#)	12	0.001175	0	0
43	(C.D.F.G#)	(D.G.B)	(D.G.B)	(C.D#.G)	12	0.001175	0	0
44	(C.D#.G)	(C.D#.G)	(C.D.F.G#)	(C.D#.G)	12	0.001175	1	9.6525e-005
45	(C.D#.G)	(C.D.F.G#)	(D.G.B)	(C.E.G)	12	0.001175	0	0
46	(C.D#.G.A#)	(C.F.A)	(D.F.A#)	(D#.G.A#)	12	0.001175	0	0
47	(C.D#.G)	(C.D.G)	(C.D#.G)	(C.D#.G)	12	0.001175	0	0
48	(C.D#.G)	(C.D#.G)	(C.D.F.G#)	(D.G.B)	12	0.001175	2	0.00019305
49	(C.D#.G)	(C.D#.G)	(C.D.G)	(D.G.B)	11	0.0010771	1	9.6525e-005
50	(D.G.B)	(C.D#.G)	(D.G.B)	(C.D#.G#)	11	0.0010771	2	0.00019305

50 most frequent 4-grams in major using harmonic approximation

					f	*p*	*rf*	*rp*
1	(C.E.G)	(D.G.B)	(C.E.G)	(C.E.G)	76	0.006947	59	0.005337
2	(D.G.B)	(C.E.G)	(C.E.G)	(C.E.G)	69	0.0063071	40	0.0036183
3	(C.E.G)	(C.D.F.A)	(D.G.B)	(C.E.G)	66	0.0060329	2	0.00018091
4	(C.E.G)	(C.E.G)	(D.G.B)	(C.E.G)	64	0.0058501	59	0.005337
5	(D.G.B)	(C.E.G)	(C.F.A)	(C.E.G)	57	0.0052102	6	0.00054274
6	(D.G.B)	(C.E.G)	(C.E.G)	(C.F.A)	56	0.0051188	19	0.0017187
7	(C.F.A)	(C.E.G)	(D.G.B)	(C.E.G)	52	0.0047532	13	0.0011759
8	(D.G.B)	(C.E.G)	(D.G.B)	(C.E.G)	50	0.0045704	23	0.0020805
9	(C.E.G)	(D.G.B)	(D.G.B)	(C.E.G)	49	0.004479	29	0.0026232
10	(C.D.F.A)	(D.G.B)	(C.E.G)	(C.E.G)	47	0.0042962	3	0.00027137
11	(D.G.B)	(C.E.G)	(C.E.G)	(D.G.B)	45	0.0041133	29	0.0026232
12	(C.E.G)	(C.F.A)	(C.E.G)	(D.G.B)	43	0.0039305	10	0.00090457
13	(C.E.G)	(D.G.B)	(C.E.G)	(C.F.A)	39	0.0035649	11	0.00099502
14	(C.E.G)	(C.E.G)	(C.E.G)	(D.G.B)	39	0.0035649	34	0.0030755
15	(C.E.G)	(D.G.B)	(C.E.G)	(D.G.B)	38	0.0034735	27	0.0024423
16	(C.E.A)	(C.D.F.A)	(D.G.B)	(C.E.G)	37	0.0033821	0	0
17	(C.E.G)	(C.E.G)	(C.E.G)	(C.F.A)	35	0.0031993	26	0.0023519
18	(D.G.B)	(C.E.G)	(C.E.G)	(C.E.A)	35	0.0031993	20	0.0018091
19	(C.E.G)	(C.E.G)	(C.E.G)	(C.E.G)	33	0.0030165	60	0.0054274
20	(D.G.B)	(D.G.B)	(C.E.G)	(C.E.G)	33	0.0030165	37	0.0033469
21	(C.E.G)	(C.E.G)	(C.F.A)	(D.G.B)	32	0.002925	17	0.0015378
22	(C.E.G)	(C.F.A)	(D.G.B)	(C.E.G)	31	0.0028336	11	0.00099502
23	(D.F.G.B)	(C.E.G)	(D.G.B)	(C.E.G)	31	0.0028336	6	0.00054274
24	(D.G.B)	(C.E.G)	(C.F.A)	(D.G.B)	30	0.0027422	12	0.0010855
25	(C.E.G)	(C.D.G)	(D.G.B)	(C.E.G)	30	0.0027422	1	9.0457e-005
26	(C.E.G)	(C.E.G)	(C.F.A)	(C.E.G)	27	0.002468	22	0.00199
27	(D.G.B)	(C.E.A)	(D.G.B)	(C.E.G)	27	0.002468	11	0.00099502
28	(D.F#.A)	(D.G.B)	(D.G.B)	(C.E.G)	27	0.002468	2	0.00018091
29	(C.E.G)	(C.F.A)	(D.F.B)	(C.E.G)	26	0.0023766	2	0.00018091
30	(E.G#.B)	(C.E.A)	(D.G.B)	(C.E.G)	25	0.0022852	1	9.0457e-005
31	(C.E.G)	(D.F.A)	(D.G.B)	(C.E.G)	24	0.0021938	5	0.00045228
32	(C.E.A)	(D.G.B)	(C.E.G)	(C.E.G)	24	0.0021938	9	0.00081411
33	(D.G.B)	(C.E.G)	(D.G.B)	(D.G.B)	24	0.0021938	17	0.0015378
34	(C.E.G)	(C.E.G)	(C.F.A)	(D.F.A)	23	0.0021024	3	0.00027137
35	(D.G.B)	(C.E.G.A)	(D.F#.A)	(D.G.B)	22	0.002011	0	0
36	(C.E.G)	(C.E.G)	(D.G.B)	(C.E.A)	22	0.002011	10	0.00090457
37	(D.G.B)	(D.G.B)	(C.E.G)	(D.G.B)	22	0.002011	28	0.0025328
38	(C.E.G)	(C.E.G)	(D.F.G.B)	(C.E.G)	22	0.002011	9	0.00081411
39	(C.E.G)	(D.G.B)	(C.E.A)	(D.G.B)	22	0.002011	9	0.00081411
40	(C.E.A)	(D.G.B)	(C.E.G)	(C.F.A)	21	0.0019196	5	0.00045228
41	(C.F.A)	(D.G.B)	(C.F.A)	(C.E.G)	21	0.0019196	5	0.00045228
42	(C.F.A)	(C.E.G)	(C.F.A)	(C.E.G)	20	0.0018282	1	9.0457e-005
43	(C.E.G)	(C.F.A)	(D.G.B)	(C.F.A)	20	0.0018282	4	0.00036183
44	(C.E.G)	(C.F.A)	(C.E.G)	(C.F.A)	19	0.0017367	3	0.00027137
45	(D.G.B)	(D.G.B)	(C.E.G)	(C.F.A)	19	0.0017367	7	0.0006332
46	(C.E.G.A)	(D.F#.A)	(D.G.B)	(C.E.G)	19	0.0017367	0	0
47	(C.F.A)	(C.E.G)	(D.F.G.B)	(C.E.G)	19	0.0017367	0	0
48	(C.E.A)	(D.G.B)	(C.E.G)	(D.G.B)	19	0.0017367	5	0.00045228
49	(C.E.G)	(C.E.G)	(D.G.B)	(D.G.B)	19	0.0017367	29	0.0026232
50	(D.G.B)	(C.E.G)	(D.F.G.B)	(C.E.G)	18	0.0016453	7	0.0006332

50 most frequent 4-grams in minor using harmonic approximation

					f	***p***	***rf***	***rp***
1	(C.D#.G)	(C.D#.G)	(D.G.B)	(C.D#.G)	67	0.0065603	12	0.0011583
2	(C.D#.G)	(D.G.B)	(C.D#.G)	(C.D#.G)	64	0.0062665	12	0.0011583
3	(D.G.B)	(C.D#.G)	(C.D#.G)	(C.D#.G)	44	0.0043082	9	0.00086873
4	(D.G.B)	(C.D#.G)	(D.G.B)	(C.D#.G)	43	0.0042103	4	0.0003861
5	(D.G.B)	(C.D#.G)	(D.F.A#)	(D#.G.A#)	42	0.0041124	5	0.00048263
6	(C.D#.G)	(C.D.F.G#)	(D.G.B)	(C.D#.G)	41	0.0040145	2	0.00019305
7	(C.D.F.G#)	(D.G.B)	(C.D#.G)	(C.D#.G)	40	0.0039166	1	9.6525e-005
8	(D.F.A#)	(D#.G.A#)	(D#.G.A#)	(D.F.A#)	36	0.0035249	0	0
9	(C.D#.G)	(C.F.G#)	(D.G.B)	(C.D#.G)	34	0.0033291	3	0.00028958
10	(D#.G.A#)	(C.D#.G#)	(D#.G.A#)	(D#.G.A#)	30	0.0029374	2	0.00019305
11	(C.D#.G)	(D.G.B)	(C.D#.G)	(D.G.B)	30	0.0029374	5	0.00048263
12	(D#.G.A#)	(D.F.A#)	(D#.G.A#)	(D#.G.A#)	29	0.0028395	3	0.00028958
13	(C.F.G#)	(C.D#.G)	(D.G.B)	(C.D#.G)	27	0.0026437	5	0.00048263
14	(D#.G.A#)	(D#.G.A#)	(D.F.A#)	(D#.G.A#)	26	0.0025458	1	9.6525e-005
15	(D.G.B)	(C.D#.G)	(C.D#.G)	(D.G.B)	26	0.0025458	7	0.00067568
16	(D#.G.A#)	(D#.G.A#)	(D#.G.A#)	(D.F.A#)	25	0.0024479	0	0
17	(D.G.B)	(C.D#.G)	(C.F.G#)	(D.G.B)	25	0.0024479	6	0.00057915
18	(D#.G.A#)	(C.D#.F.G#)	(D.F.A#)	(D#.G.A#)	24	0.0023499	0	0
19	(C.F.G#)	(D.G.B)	(D.G.B)	(C.D#.G)	24	0.0023499	2	0.00019305
20	(C.D#.G)	(D.G.B)	(C.D#.G)	(C.F.G#)	24	0.0023499	8	0.0007722
21	(D.F.A#)	(D#.G.A#)	(D#.G.A#)	(D#.G.A#)	23	0.002252	2	0.00019305
22	(C.D#.G)	(D.G.B)	(D.G.B)	(C.D#.G)	23	0.002252	12	0.0011583
23	(C.D#.G)	(D#.G.A#)	(C.D#.G#)	(D#.G.A#)	22	0.0021541	1	9.6525e-005
24	(C.D#.G)	(C.D#.G)	(C.D#.G)	(D.G.B)	22	0.0021541	16	0.0015444
25	(C.D#.F.G#)	(D.F.A#)	(D#.G.A#)	(D#.G.A#)	22	0.0021541	0	0
26	(D.G.B)	(D.G.B)	(C.D#.G)	(C.D#.G)	21	0.0020562	10	0.00096525
27	(C.D.G)	(D.G.B)	(C.D#.G)	(C.D#.G)	20	0.0019583	1	9.6525e-005
28	(D#.G.A#)	(C.F.A)	(D.F.A#)	(D#.G.A#)	20	0.0019583	0	0
29	(C.F.G#)	(D.G.B)	(C.D#.G)	(C.D#.G)	19	0.0018604	6	0.00057915
30	(C.D#.G)	(C.D#.G)	(C.D#.G)	(C.D#.G)	18	0.0017625	15	0.0014479
31	(C.D#.G)	(C.D.G)	(D.G.B)	(C.D#.G)	18	0.0017625	2	0.00019305
32	(D.G.B)	(C.D#.G)	(C.F.A)	(D.F.A#)	18	0.0017625	2	0.00019305
33	(D.G.B)	(C.D#.G)	(C.D#.G)	(C.D#.G#)	18	0.0017625	4	0.0003861
34	(C.D#.G)	(D.F.A#)	(D#.G.A#)	(D#.G.A#)	18	0.0017625	2	0.00019305
35	(D.G.B)	(D.G.B)	(C.D#.G)	(D.F.A#)	18	0.0017625	6	0.00057915
36	(D.F.A#)	(D#.G.A#)	(D.F.A#)	(D#.G.A#)	17	0.0016645	0	0
37	(C.D#.G#)	(D#.G.A#)	(D#.G.A#)	(D#.G.A#)	17	0.0016645	0	0
38	(C.D#.G)	(C.F.G#)	(C.D.F.G#)	(D.G.B)	17	0.0016645	0	0
39	(D.F.A#)	(D#.G.A#)	(D#.G.A#)	(C.D#.G#)	17	0.0016645	2	0.00019305
40	(D.G.B)	(C.D#.G)	(C.D#.G)	(D#.G.A#)	17	0.0016645	7	0.00067568
41	(C.D#.G)	(D.G.B)	(C.D#.G)	(D.F.A#)	16	0.0015666	3	0.00028958
42	(C.D#.G)	(C.D#.G)	(D.F.A#)	(D#.G.A#)	16	0.0015666	3	0.00028958
43	(C.D#.G)	(C.D#.F.G#)	(D.F.A#)	(D#.G.A#)	16	0.0015666	0	0
44	(C.D#.G)	(D.F.A#)	(D#.G.A#)	(C.D#.G)	16	0.0015666	3	0.00028958
45	(D.G.B)	(C.D#.G)	(C.D#.G)	(C.F.G#)	15	0.0014687	4	0.0003861
46	(C.D#.G)	(C.D#.G)	(D.F.G.B)	(C.D#.G)	15	0.0014687	5	0.00048263
47	(C.D#.G)	(C.D.F.G#)	(D.G.B)	(C.E.G)	15	0.0014687	0	0
48	(C.D#.G)	(D.F.A#)	(D#.G.A#)	(C.D#.G#)	14	0.0013708	1	9.6525e-005
49	(D.G.B)	(C.D#.G)	(C.D#.G)	(C.F.A)	14	0.0013708	4	0.0003861
50	(C.D#.G)	(D.F.A#)	(D#.G.A#)	(D.F.A#)	14	0.0013708	2	0.00019305

50 most frequent 4-grams in major within a random corpus

					f	*p*
1	(C.E.G)	(C.E.G)	(C.E.G)	(C.E.G)	21	0.0011234
2	(C.E.G)	(D.G.B)	(C.E.G)	(C.E.G)	18	0.00096293
3	(C.E.G)	(C.E.G)	(C.E.G)	(D.G.B)	18	0.00096293
4	(C.E.G)	(C.E.G)	(D.G.B)	(C.E.G)	17	0.00090943
5	(C.E.G)	(D.G.B)	(C.E.G)	(D.G.B)	14	0.00074894
6	(C.F.A)	(C.E.G)	(C.E.G)	(C.E.G)	14	0.00074894
7	(D.G.B)	(D.G.B)	(C.E.G)	(C.E.G)	12	0.00064195
8	(C.E.G)	(C.E.G)	(C.F.A)	(C.E.G)	12	0.00064195
9	(C.E.G)	(C.E.G)	(D.G.B)	(D.G.B)	12	0.00064195
10	(C.E.G)	(D.G.B)	(D.G.B)	(C.E.G)	11	0.00058846
11	(D.G.B)	(C.E.G)	(C.E.G)	(D.G.B)	11	0.00058846
12	(D.G.B)	(C.E.G)	(D.G.B)	(C.E.G)	10	0.00053496
13	(D.G.B)	(D.G.B)	(C.E.G)	(C.F.A)	10	0.00053496
14	(D.G.B)	(C.E.G)	(C.E.G)	(C.E.G)	9	0.00048146
15	(D.G.B)	(D.G.B)	(D.G.B)	(C.E.G)	8	0.00042797
16	(C.E.G)	(D.F.A)	(D.G.B)	(C.E.G)	8	0.00042797
17	(C.E.G)	(C.F.A)	(D.G.B)	(C.E.G)	8	0.00042797
18	(C.E.G)	(D.G.B)	(C.E.A)	(C.E.G)	8	0.00042797
19	(D.G.B)	(C.E.G)	(C.E.G)	(C.E.A)	8	0.00042797
20	(C.F.A)	(C.E.G)	(D.G.B)	(C.E.G)	7	0.00037447
21	(C.E.G)	(C.E.G)	(C.E.G)	(C.F.A)	7	0.00037447
22	(C.E.G)	(C.E.G)	(C.E.A)	(D.G.B)	7	0.00037447
23	(C.E.G)	(C.E.G)	(D.G.B)	(C.E.A)	7	0.00037447
24	(D.F.G.B)	(C.E.G)	(D.G.B)	(C.E.G)	6	0.00032098
25	(C.E.G)	(D.G.B)	(D.G.B)	(D.G.B)	6	0.00032098
26	(C.E.G)	(C.E.G)	(D.G.B)	(C.F.A)	6	0.00032098
27	(C.E.G)	(D.F.G.B)	(C.E.G)	(C.E.G)	6	0.00032098
28	(D.G.B)	(C.E.G)	(D.F.G.B)	(D.G.B)	6	0.00032098
29	(C.E.G)	(C.E.G)	(C.F.A)	(C.F.A)	6	0.00032098
30	(C.E.G)	(C.F.A)	(C.E.G)	(C.E.G)	6	0.00032098
31	(C.E.G)	(C.E.G)	(C.E.A)	(C.E.G)	6	0.00032098
32	(E.G.B)	(C.E.G)	(C.E.G)	(C.E.G)	6	0.00032098
33	(C.E.G)	(C.E.G)	(D.G.B)	(D.F.G.B)	6	0.00032098
34	(C.E.G)	(C.F.A)	(C.E.G)	(C.E.A)	6	0.00032098
35	(C.F.A)	(C.E.G)	(D.G.B)	(D.G.B)	5	0.00026748
36	(D.G.B)	(C.E.G)	(D.G.B)	(C.E.A)	5	0.00026748
37	(D.F.G.B)	(C.E.G)	(C.E.G)	(C.E.G)	5	0.00026748
38	(D.G.B)	(C.F.A)	(C.E.G)	(D.G.B)	5	0.00026748
39	(C.F.A)	(C.E.G)	(C.E.A)	(C.E.G)	5	0.00026748
40	(C.E.G)	(D.G.B)	(C.F.A)	(C.E.G)	5	0.00026748
41	(C.E.G)	(C.E.A)	(C.E.G)	(C.E.G)	5	0.00026748
42	(D.F.G.B)	(D.G.B)	(C.E.G)	(D.G.B)	5	0.00026748
43	(C.E.A)	(D.G.B)	(D.G.B)	(C.E.G)	5	0.00026748
44	(D.G.B)	(E.G.B)	(C.E.G)	(C.E.G)	5	0.00026748
45	(C.E.G)	(D.G.B)	(E.G.B)	(C.E.G)	5	0.00026748
46	(C.E.G)	(C.E.G)	(D.F.A)	(D.G.B)	5	0.00026748
47	(D.G.B)	(C.E.G)	(C.F.A)	(C.E.G)	5	0.00026748
48	(C.E.G)	(C.E.A)	(D.G.B)	(C.E.G)	5	0.00026748
49	(C.F.A)	(D.G.B)	(C.E.G)	(C.E.G)	5	0.00026748
50	(D.G.B)	(C.E.A)	(C.E.G)	(D.G.B)	5	0.00026748

50 most frequent 4-grams in minor within a random corpus

					f	***p***
1	(D#.G.A#)	(C.D#.G)	(C.D#.G)	(C.D#.G)	6	0.00034874
2	(C.D#.G)	(C.D#.G)	(C.D#.G)	(D.G.B)	6	0.00034874
3	(C.D#.G)	(C.D#.G)	(C.D#.G)	(C.D#.G)	6	0.00034874
4	(C.D#.G)	(D.G.B)	(D#.G.A#)	(C.D#.G)	6	0.00034874
5	(C.D#.G)	(D.F.A#)	(C.D#.G)	(C.D#.G)	6	0.00034874
6	(D.G.B)	(D#.G.A#)	(C.D#.G)	(C.D#.G)	5	0.00029061
7	(D#.G.A#)	(C.D#.G)	(C.D#.G)	(D.G.B)	5	0.00029061
8	(C.D#.G)	(C.D#.G)	(D.F.A#)	(C.D#.G)	5	0.00029061
9	(D.G.A#)	(D#.G.A#)	(C.D#.G)	(C.D#.G)	4	0.00023249
10	(C.D#.G)	(C.D#.G)	(C.D#.G)	(C.F.G#)	4	0.00023249
11	(D.F.A#)	(C.F.G#)	(C.D#.G)	(C.D#.G)	4	0.00023249
12	(C.D#.G)	(D.F.A#)	(D.F.A#)	(C.D#.G)	4	0.00023249
13	(C.D#.G)	(C.D#.G)	(D.G.B)	(C.D#.G)	4	0.00023249
14	(D#.G.A#)	(C.D#.G)	(C.D#.G)	(C.F.G#)	4	0.00023249
15	(D#.G.A#)	(D#.G.A#)	(C.D#.G)	(C.D#.G)	4	0.00023249
16	(C.D#.G)	(C.D#.F.G#)	(C.D#.G)	(D.F.A#)	4	0.00023249
17	(D#.G.A#)	(C.D#.G)	(D.G.B)	(C.D#.G)	4	0.00023249
18	(D.G.B)	(C.D#.G)	(D.F.A#)	(C.D#.G)	4	0.00023249
19	(C.D#.G)	(C.D#.G)	(D.G.B)	(D.F.G.B)	4	0.00023249
20	(C.D#.G)	(D#.G.A#)	(D#.G.A#)	(D#.G.A#)	4	0.00023249
21	(C.D#.G)	(C.D#.G)	(C.D#.G)	(D.F.A#)	4	0.00023249
22	(C.D#.G)	(D#.G.A#)	(D.G.B)	(C.D#.G)	4	0.00023249
23	(C.D#.G)	(D#.G.A#)	(C.D#.G)	(C.D#.G)	4	0.00023249
24	(D#.G.A#)	(D.G.B)	(D.G.B)	(C.D#.G)	4	0.00023249
25	(C.D#.G)	(C.D#.G)	(C.D.G)	(C.D#.G)	4	0.00023249
26	(C.D#.G)	(C.D#.G.A#)	(C.D#.G)	(D#.G.A#)	3	0.00017437
27	(D.F.A#)	(C.D#.G)	(C.D#.G.A#)	(C.D#.G)	3	0.00017437
28	(D.F.A#)	(C.D#.G)	(C.D#.G)	(C.D#.G.A#)	3	0.00017437
29	(C.D#.F.G#)	(C.D#.G)	(C.D#.G)	(D.D#.G.A#)	3	0.00017437
30	(D#.G.A#)	(C.F.G#)	(D.F.A#)	(C.D#.G)	3	0.00017437
31	(D.G.B)	(D#.G.A#)	(C.D#.G)	(D.G.B)	3	0.00017437
32	(C.D#.G)	(D.F.A#)	(D.G.B)	(C.D.F.G)	3	0.00017437
33	(C.D.F.G#)	(D#.G.A#)	(C.D#.G)	(C.D#.G)	3	0.00017437
34	(D.G.A#)	(D.F.A#)	(C.D#.G)	(C.D#.G)	3	0.00017437
35	(C.D#.G)	(D.D#.G.A#)	(D.F.A#)	(C.D#.G)	3	0.00017437
36	(C.F.G#)	(C.D#.G)	(D#.G.A#)	(D#.G.A#)	3	0.00017437
37	(D#.G.A#)	(C.D#.G)	(D#.G.A#)	(D#.G.A#)	3	0.00017437
38	(C.F.A)	(C.D#.G)	(C.D#.G)	(C.D#.G)	3	0.00017437
39	(C.F.G#)	(C.D#.G)	(D.G.A#)	(C.D#.G)	3	0.00017437
40	(C.D#.G)	(D.G.B)	(D.F.A#)	(C.D#.G)	3	0.00017437
41	(D.G.B)	(D.F.A#)	(C.D#.G#)	(C.D#.G)	3	0.00017437
42	(D.G.B)	(D.F.A#)	(C.D#.G)	(D#.G.A#)	3	0.00017437
43	(D.F.A#)	(D.F.A#)	(D.F.A#)	(C.D#.G)	3	0.00017437
44	(C.D#.G)	(C.D#.G)	(C.E.G)	(C.D#.G)	3	0.00017437
45	(C.D#.G)	(D#.G.A#)	(C.D#.G)	(D.G.A#)	3	0.00017437
46	(C.D#.G)	(C.F.G#)	(C.F.G#)	(C.D#.G)	3	0.00017437
47	(D.G.B)	(C.F.G#)	(C.D#.G)	(D#.G.A#)	3	0.00017437
48	(C.D#.G)	(D#.G.A#)	(D.G.B)	(D#.G.A#)	3	0.00017437
49	(C.D.D#.G)	(D.F.A#)	(C.D#.G)	(C.D#.G)	3	0.00017437
50	(C.D#.G)	(C.D#.G)	(C.D#.G)	(D.G.A#)	3	0.00017437

Appendix B

Tables and Figures

Figure 3.1. Different methods of segmentation applied to the excerpt at the top:
(a) dense segmentation (b) metric segmentation (c) harmonic approximation
(d) maximal segmentation (e) illustration that dense segmentation does not include information on rhythm or metre.

Pc set	*dissonance rating*	*Pc set*	*dissonance rating*
(C.E.G)	2	(C.D.E.F#)	-4
(C.D#.G)	2	(C.E.F)	-4
(C.D#.F#.G#)	1	(C.C#.F)	-4
(C.D.F#)	1	(C.D#.E)	-4
(C.F)	0	(C.C#.E)	-4
(C.E)	0	(C.C#)	-4
(C.D#)	0	(C.D#.F#.G)	-5
(C)	0	(C.E.F.G)	-5
(C.D#.F.G#)	-1	(C.C#.E.G)	-5
(C.D.G)	-1	(C.D.D#.G)	-5
(C.D#.F#)	-1	(C.F.F#)	-5
(C.F#)	-1	(C.C#.F#)	-5
(C.D#.F)	-1	(C.D.D#)	-5
(C.D.F)	-1	(C.C#.D#)	-5
(C.D)	-1	(C.E.F#.G)	-6
(C.D#.F#.A)	-2	(C.D.F#.G)	-6
(C.D.F.G#)	-2	(C.C#.F.G)	-6
(C.D#.F.G)	-2	(C.C#.D#.G)	-6
(C.D.F.G)	-2	(C.D#.F.F#)	-6
(C.D.E.G)	-2	(C.D#.E.F#)	-6
(C.E.F#)	-2	(C.D.D#.F#)	-6
(C.D.E)	-2	(C.C#.D#.F#)	-6
(C.D.E.G#)	-3	(C.D.E.F)	-6
(C.E.G#)	-3	(C.D.D#.F)	-6
(C.D.F#.G#)	-4	(C.C#.D#.F)	-6
(C.C#.F.G#)	-4	(C.C#.E.F)	-8
(C.D#.E.G#)	-4	(C.C#.F.F#)	-9
(C.C#.E.G#)	-4	(C.C#.D#.E)	-9
(C.D#.E.G)	-4		

Table 3.1. Dissonance ratings for all different pc set genera

	Major				Minor			
Rank	Pc set	diat. scale degree	Frequency	Relative frequency	Pc set	diat. scale degree	Frequency	Relative frequency
1	(C.E.G)	I	3446	0.18451	(C.D#.G)	I	2526	0.13879
2	(D.G.B)	V	2254	0.12069	(D#.G.A#)	III	1458	0.08011
3	(C.F.A)	IV	1146	0.061362	(D.G.B)	V	1317	0.072363
4	(C.E.A)	VI	1025	0.054883	(D.F.A#)	VII	1047	0.057527
5	(D.F.G.B)	V7	882	0.047226	(C.F.G#)	IV	877	0.048187
6	(D.F.A)	II	698	0.037374	(D.F.G.B)	V7	653	0.035879
7	(E.G.B)	III	433	0.023185	(C.D#.G#)	VI	621	0.034121
8	(C.D.F.A)		392	0.02099	(D.G.A#)		559	0.030714
9	(D.F.B)	VII	374	0.020026	(C.D.F.G#)		427	0.023462
10	(C.E.G.A)		358	0.019169	(C.F.A)		364	0.02
11	(C.D.F#.A)		357	0.019115	(D.F.G#.A#)		356	0.01956
12	(D.F#.A)		353	0.018901	(C.E.G)		356	0.01956
13	(C.D.G)		334	0.017884	(D.F.G#)	II	337	0.018516
14	(C.E.F.A)		313	0.016759	(C.D#.G.A#)		332	0.018242
15	(C.E.G.B)		295	0.015796	(C.D#.F.G#)		298	0.016374
16	(C.D.E.G)		267	0.014296	(C.D.G)		297	0.016319
17	(E.G#.B)		210	0.011244	(C.D#.F.A)		274	0.015055
18	(D.E.G.B)		204	0.010923	(C.D.D#.G)		230	0.012637
19	(C.D.F.G)		192	0.010281	(D.F.B)		229	0.012582
20	(C.E.G.A#)		178	0.0095309	(D.D#.G.A#)		177	0.0097253

Table 5.1. Top 20 pc sets for major and minor

Triads in minor		
	in natural minor	*other quality*
I	minor	none
II	diminished	minor (with raised $\hat{6}$) - infrequent
III	major	augmented (with raised $\hat{7}$) - infrequent
IV	minor	major (with raised $\hat{6}$) - infrequent
V	minor	major (with raised $\hat{7}$) - very frequent
VI	major	diminished (with raised $\hat{6}$) - infrequent
VII	major	diminished (with raised $\hat{7}$) - very frequent

Table 5.2. from Aldwell & Schachter (1989), p.51

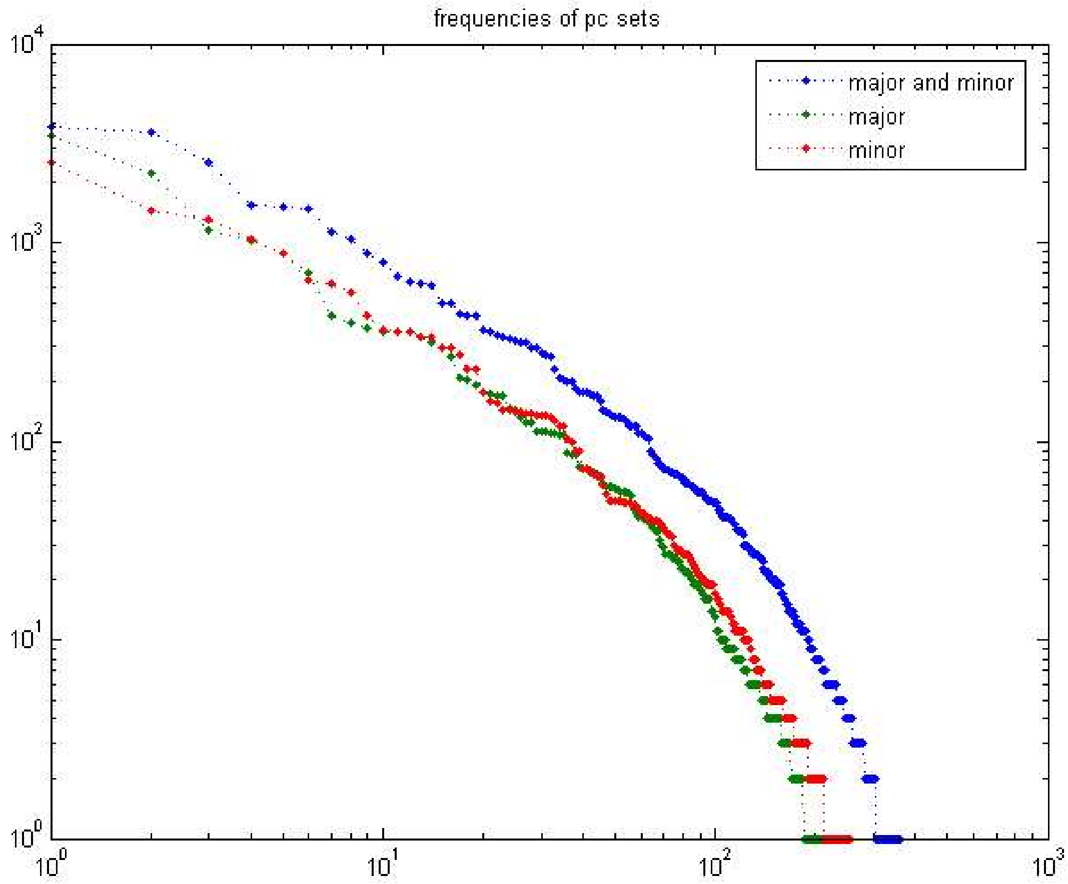


Diagram 5.1, plotting frequencies (y-axis) of all pc sets found against their rank

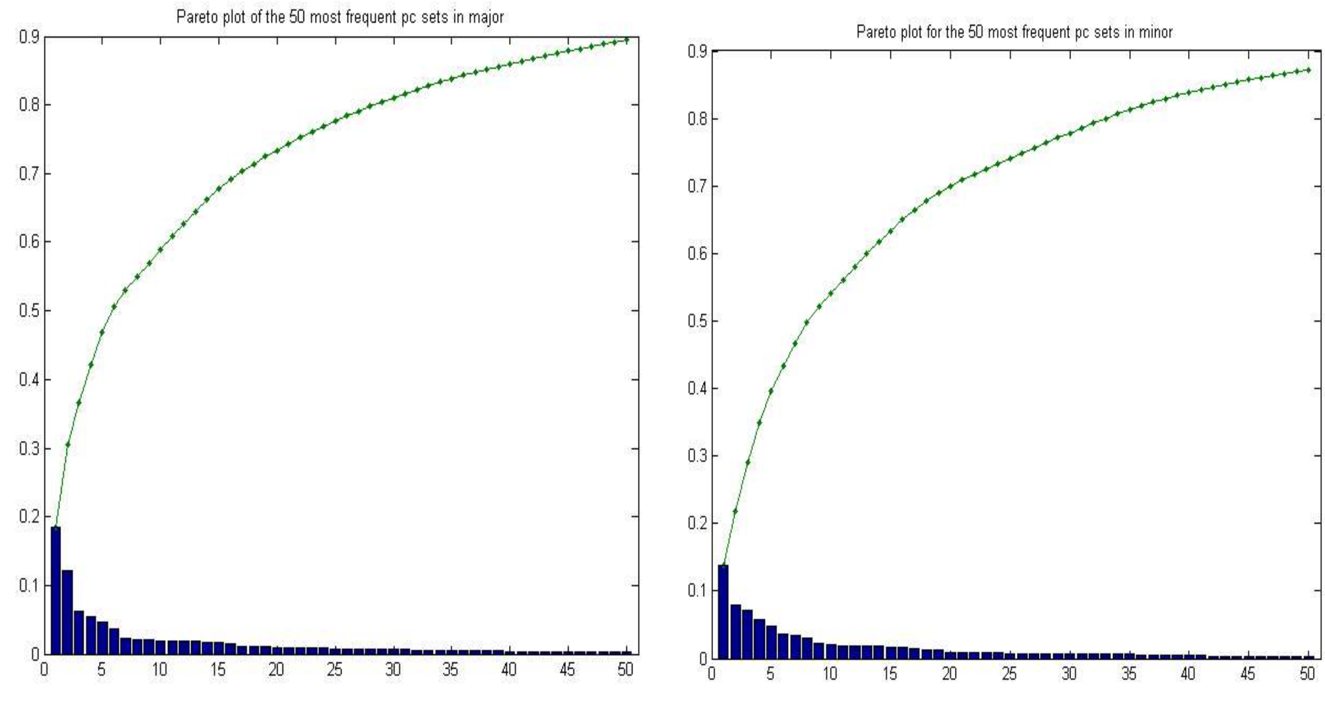


Diagram 5.2 Pareto plots combining rank and (cumulative) frequencies for the 50 most frequent pc sets.

	Pc set genera in major					Pc set genera in minor				
	pc set	Frequency	Relative frequency	Instances	Pc set rating	Pc set	Frequency	Relative frequency	Instances	Pc set rating
1	(C.E.G)	7655	0.40988	9	2	(C.E.G)	5331	0.29291	10	2
2	(C.D#.G)	2325	0.12449	8	2	(C.D#.G)	4122	0.22648	11	2
3	(C.D#.F#.G#)	1654	0.088563	7	1	(C.D#.F#.G#)	1612	0.088571	9	1
4	(C.D.F#)	220	0.01178	8	1	(C.D.F#)	194	0.010659	8	1
5	(C.F)	177	0.0094774	7	0	(C.F)	152	0.0083516	7	0
6	(C.E)	277	0.014832	7	0	(C.E)	154	0.0084615	8	0
7	(C.D#)	125	0.0066931	6	0	(C.D#)	161	0.0088462	5	0
8	(C)	21	0.0011244	6	0	(C)	5	0.00027473	3	0
9	(C.D#.F.G#)	1024	0.05483	7	-1	(C.D#.F.G#)	844	0.046374	8	-1
10	(C.D.G)	740	0.039623	7	-1	(C.D.G)	807	0.044341	8	-1
11	(C.D#.F#)	709	0.037963	12	-1	(C.D#.F#)	895	0.049176	12	-1
12	(C.D#.F)	194	0.010388	5	-1	(C.F#)	3	0.00016484	2	-1
13	(C.D.F)	413	0.022114	7	-1	(C.D#.F)	172	0.0094505	6	-1
14	(C.D)	8	0.00042836	3	-1	(C.D.F)	308	0.016923	6	-1
15	(C.D#.F#.A)	99	0.0053009	3	-2	(C.D)	15	0.00082418	5	-1
16	(C.D.F.G#)	308	0.016492	7	-2	(C.D#.F#.A)	270	0.014835	3	-2
17	(C.D#.F.G)	161	0.0086207	7	-2	(C.D.F.G#)	615	0.033791	6	-2
18	(C.D.F.G)	346	0.018526	7	-2	(C.D#.F.G)	231	0.012692	6	-2
19	(C.D.E.G)	481	0.025755	7	-2	(C.D.F.G)	297	0.016319	6	-2
20	(C.E.F#)	34	0.0018205	4	-2	(C.D.E.G)	233	0.012802	7	-2
21	(C.D.E)	17	0.00091026	3	-2	(C.E.F#)	45	0.0024725	6	-2
22	(C.D.E.G#)	5	0.00026772	3	-3	(C.D.E)	19	0.001044	4	-2
23	(C.E.G#)	19	0.0010173	4	-3	(C.D.E.G#)	13	0.00071429	3	-3
24	(C.C#.F.G#)	645	0.034536	4	-4	(C.E.G#)	98	0.0053846	3	-3
25	(C.D#.E.G#)	22	0.001178	4	-4	(C.D.F#.G#)	3	0.00016484	1	-4
26	(C.C#.E.G#)	7	0.00037481	2	-4	(C.C#.F.G#)	365	0.020055	6	-4
27	(C.D#.E.G)	2	0.00010709	2	-4	(C.D#.E.G#)	62	0.0034066	3	-4
28	(C.D.E.F#)	13	0.00069608	3	-4	(C.C#.E.G#)	38	0.0020879	4	-4
29	(C.E.F)	85	0.0045513	5	-4	(C.D#.E.G)	3	0.00016484	2	-4
30	(C.C#.F)	59	0.0031591	4	-4	(C.D.E.F#)	11	0.0006044	2	-4
31	(C.D#.E)	1	5.3545e-005	1	-4	(C.E.F)	50	0.0027473	6	-4
32	(C.C#.E)	1	5.3545e-005	1	-4	(C.C#.F)	132	0.0072527	4	-4
33	(C.C#)	1	5.3545e-005	1	-4	(C.D#.E)	4	0.00021978	2	-4
34	(C.D#.F#.G)	6	0.00032127	4	-5	(C.C#.E)	4	0.00021978	1	-4
35	(C.E.F.G)	265	0.014189	7	-5	(C.D#.F#.G)	12	0.00065934	2	-5
36	(C.C#.E.G)	3	0.00016063	1	-5	(C.E.F.G)	88	0.0048352	6	-5
37	(C.D.D#.G)	175	0.0093703	5	-5	(C.C#.E.G)	19	0.001044	3	-5
38	(C.F.F#)	4	0.00021418	2	-5	(C.D.D#.G)	398	0.021868	3	-5
39	(C.C#.F#)	15	0.00080317	4	-5	(C.F.F#)	6	0.00032967	3	-5
40	(C.D.D#)	6	0.00032127	2	-5	(C.C#.F#)	28	0.0015385	4	-5
41	(C.C#.D#)	10	0.00053545	4	-5	(C.D.D#)	15	0.00082418	3	-5
42	(C.E.F#.G)	71	0.0038017	4	-6	(C.C#.D#)	4	0.00021978	2	-5

43	(C.D.F#.G)	37	0.0019812	5	-6	(C.E.F#.G)	45	0.0024725	5	-6
44	(C.C#.F.G)	7	0.00037481	3	-6	(C.D.F#.G)	55	0.003022	5	-6
45	(C.C#.D#.G)	32	0.0017134	4	-6	(C.C#.F.G)	19	0.001044	3	-6
46	(C.D#.F.F#)	48	0.0025701	4	-6	(C.C#.D#.G)	26	0.0014286	3	-6
47	(C.D#.E.F#)	11	0.00058899	4	-6	(C.D#.F.F#)	43	0.0023626	4	-6
48	(C.D.D#.F#)	2	0.00010709	2	-6	(C.D#.E.F#)	24	0.0013187	4	-6
49	(C.C#.D#.F#)	28	0.0014993	2	-6	(C.D.D#.F#)	5	0.00027473	2	-6
50	(C.D.E.F)	49	0.0026237	5	-6	(C.C#.D#.F#)	41	0.0022527	5	-6
51	(C.D.D#.F)	35	0.0018741	5	-6	(C.D.E.F)	23	0.0012637	3	-6
52	(C.C#.D#.F)	20	0.0010709	2	-6	(C.D.D#.F)	26	0.0014286	4	-6
53	(C.C#.E.F)	1	5.3545e-005	1	-8	(C.C#.D#.F)	41	0.0022527	3	-6
54	(C.C#.F.F#)	2	0.00010709	2	-9	(C.C#.E.F)	2	0.00010989	1	-8
55	(C.C#.D#.E)	1	5.3545e-005	1	-9	(C.C#.F.F#)	11	0.0006044	2	-9
56						(C.C#.D#.E)	1	5.4945e-005	1	-9

Table 5.3

	Chord frequencies in samples from Bach:		Overall chord frequencies in samples from the first part of the 18th century major		Overall chord frequencies in the entire sample	
	Major	Minor	Major	Minor	Major	Minor
I	35,37	36,45	40,41	37,53	41,71	40,21
II	9,72	8,22	8,44	7,78	7,71	6,13
III	1,3	0,06	0,91	0,95	1,29	1,2
IV	8,68	9,24	9,13	9,34	8,21	9,85
V	24,44	22,02	23,24	24,45	19,49	22,57
VI	6,95	2,21	5,01	3,19	4,98	4,7
VII	3,21	3,73	4,31	4,32	2,39	3,04
I7	0	0,13	0	0,16	0,08	0,16
II7	0,84	1,98	0,77	1,47	0,64	0,77
III7	0,05	0,06	0,05	0,09	0,06	0,05
IV7	0,08	0,99	0,51	1,21	0,24	0,74
V7	7,9	9,7	6,55	7,49	12,19	9,19
VI7	0,31	0,06	0,33	0,28	0,29	0,19
VII7	0,16	1,05	0,09	0,7	0,37	0,69
Correlation	0,99	0,86	0,99	0,86	0,98	0,88

Table 5.4. Chord frequencies (in percent) counted by Budge (1943), compiled from from table XIVa, p.46-47, table IX, p.23 and table X, p.24. The last row indicates the correlations between Budge's data and the pc set frequencies found in the chorale corpus using dense segmentation.

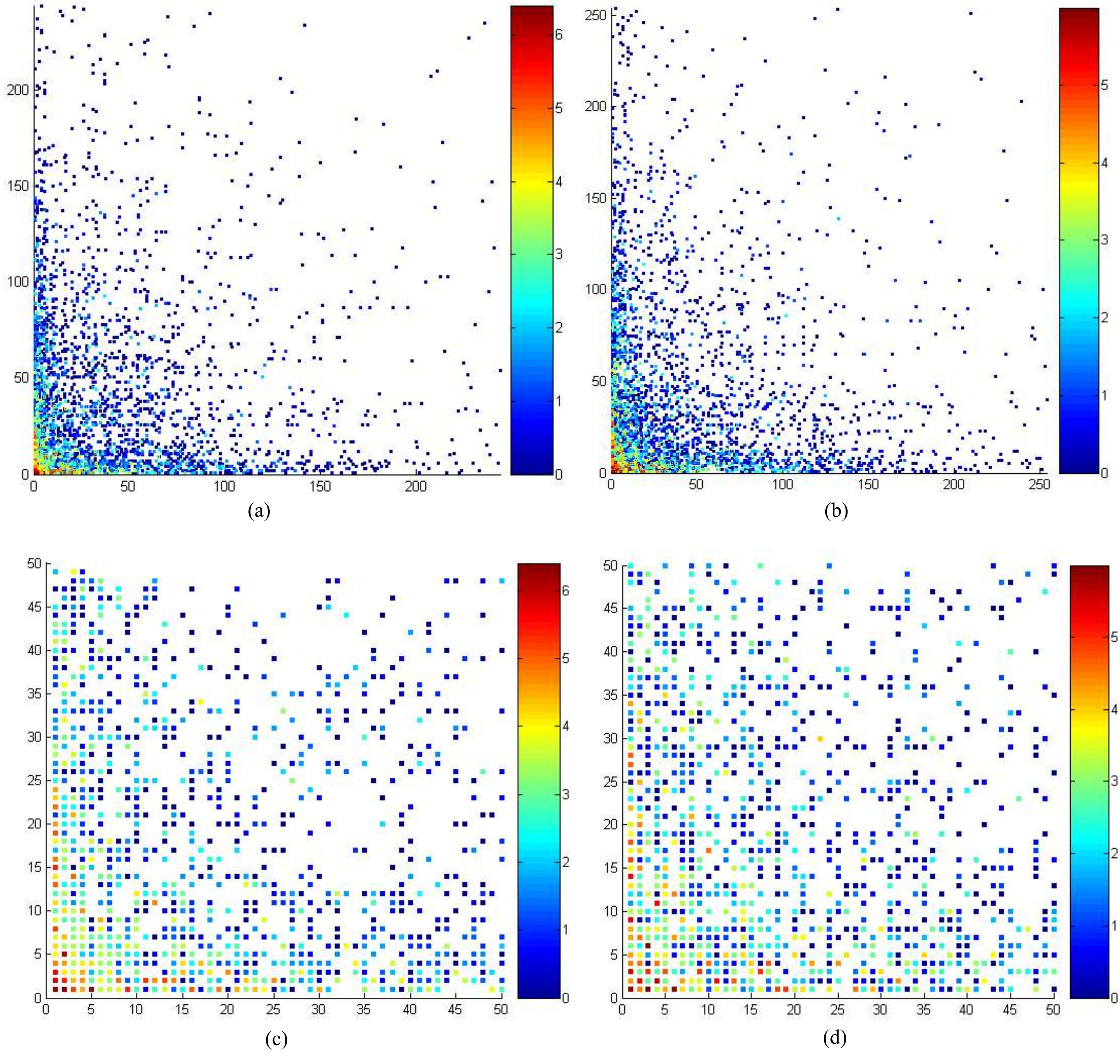


Diagram 5.3. Coloured scatter plots for pc set transitions (dense segmentation)
(a): transitions in major (b) transitions in minor
(c) transitions between the 50 most frequent pc sets in major (d) transitions between the 50 most frequent pc sets in minor

	I (C.E.G)	ii (D.F.A)	iii (E.G.B)	IV (C.F.A)	V (D.G.B)	vi (C.E.A)	vii° (D.F.B)
I (C.E.G)		5,10	4,78	5,91	5,94	5,26	4,57
ii (D.F.A)	5,69		4,00	4,76	6,10	4,97	5,41
iii (E.G.B)	5,38	4,47		4,63	5,03	4,60	4,47
IV (C.F.A)	5,94	5,00	4,22		6,00	4,35	4,79
V (D.G.B)	6,19	4,79	4,47	5,51		5,19	4,85
vi(C.E.A)	5,04	5,44	4,72	5,07	5,56		4,50
vii° (D.F.B)	5,85	4,16	4,16	4,53	5,16	4,19	

Table 5.5, Ratings of pairs of chords in major context, from a study by Bharucha & Krumhansl (1983). data reprinted in Krumhansl (1990:193)

TABLE OF USUAL ROOT PROGRESSIONS		
I is followed by IV or V,	sometimes by VI,	less often by II or III.
II is followed by V,	sometimes by IV, VI,	less often by I, III.
III is followed by VI,	sometimes by IV,	less often by I, II or V.
IV is followed by V,	sometimes by I or II,	less often by III or VI.
V is followed by I,	sometimes by VI or IV,	less often by III or II.
VI is followed by II, V,	sometimes by III, IV,	less often by I
VII is followed by III,	sometimes by I.	

Table 5.6, from Piston (1941/1978)

	(C.E.G)	(D.F.A)	(E.G.B)	(C.F.A)	(D.G.B)	(C.E.A)	(D.F.B)
(C.E.G)		132	36	474	668	191	43
(D.F.A)	116		35	11	100	59	5
(E.G.B)	47	13		73	22	52	12
(C.F.A)	351	63	31		138	29	45
(D.G.B)	1042	60	63	73		147	1
(C.E.A)	106	72	62	64	159		14
(D.F.B)	92	1	4	3	2	4	

Table 5.7a. Absolute frequencies of diatonic chord progressions in major from the Bach chorale, disregarding progressions between identical chords.

	(C.E.G)	(D.F.A)	(E.G.B)	(C.F.A)	(D.G.B)	(C.E.A)	(D.F.B)
(C.E.G)		1	1	3	3	2	
(D.F.A)	1		1	2	3	2	
(E.G.B)	1	1		2	1	3	
(C.F.A)	2	2	1		3	1	
(D.G.B)	3	1	1	2		2	
(C.E.A)	1	3	2	2	3		
(D.F.B)	2		3				

Table 5.7b. Quantification of Piston's table after Krumhansl (1990)

	(C.E.G)	(D.F.A)	(E.G.B)	(C.F.A)	(D.G.B)	(C.E.A)	(D.F.B)
(C.E.G)	0	2	2	24	24	8	0
(D.F.A)	2	0	2	8	24	8	0
(E.G.B)	2	2	0	8	2	24	0
(C.F.A)	8	8	2	0	24	2	0
(D.G.B)	24	2	2	8	0	8	0
(C.E.A)	2	24	8	8	24	0	0
(D.F.B)	8	0	24	0	0	0	0

Table 5.7c. Another quantification of Piston's table, using the values 2,8,24
(resulting from applying $a \cdot 2^a$ *to Krumhansl's quantification values 1,2,3)*

Correlations ***between***	progression frequencies in the chorale corpus (table 5.7a)	quantified progressions after Piston (table 5.7c)	ratings of chord pairs after Bharucha & Krumhansl (1983) (table 5.5)
progression frequencies in the chorale corpus (table 5.7a)	-	0.54	0.62
quantified progressions after Piston (table 5.7c)	-	-	0.54

Table 5.7d.

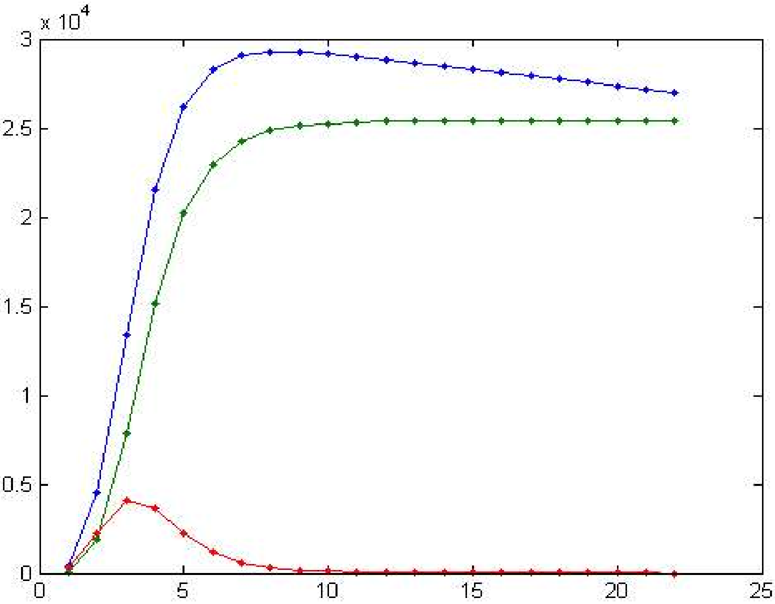


Diagram 5.4.
blue curve: number of different *n*-grams in the corpus
green curve: number of different unique *n*-grams
red curve: number of different *n*-grams which occur at least in two different chorales.

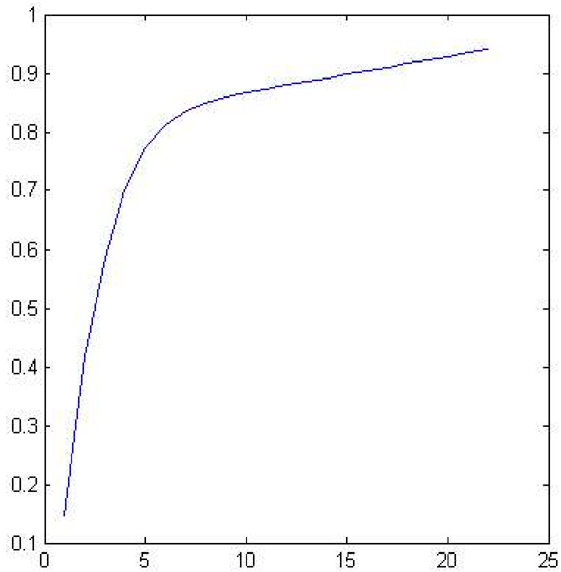
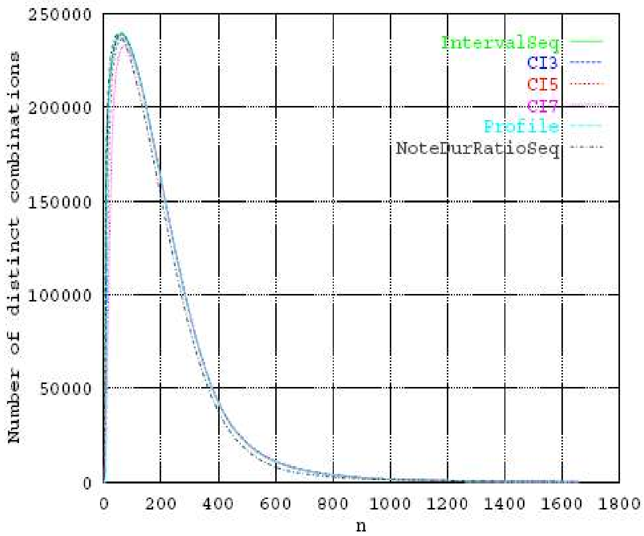


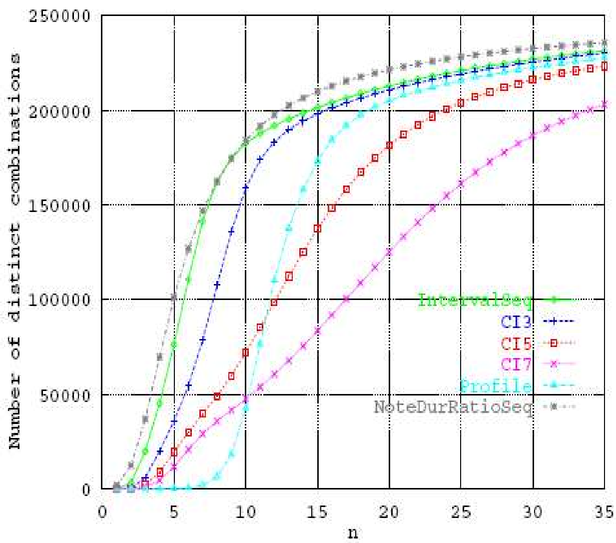
Diagram 5.5. Proportion between the number of different *n*-grams and the number of different unique *n*-grams

n	*Mean percentage*	*Standard deviation*
2	19.82%	6.41%
3	33.03%	6.81%
4	51.21%	7.82%
5	71.38%	8.84%
6	86.56%	6.20%
7	94.26%	3.95%

Table 5.8. Mean percentages of prediction of the database using n-grams (dense segmentation)



(b) Number of distinct n-grams vs. n, all features



(c) Curves zoomed in for $n \leq 35$

Diagram 5.6, reprinted from Yip & Kao (1999)

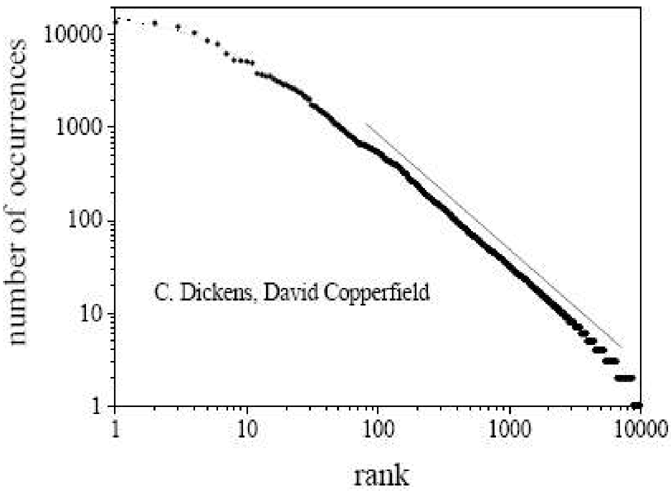


Diagram 5.7, reprinted from Zanette (2005)

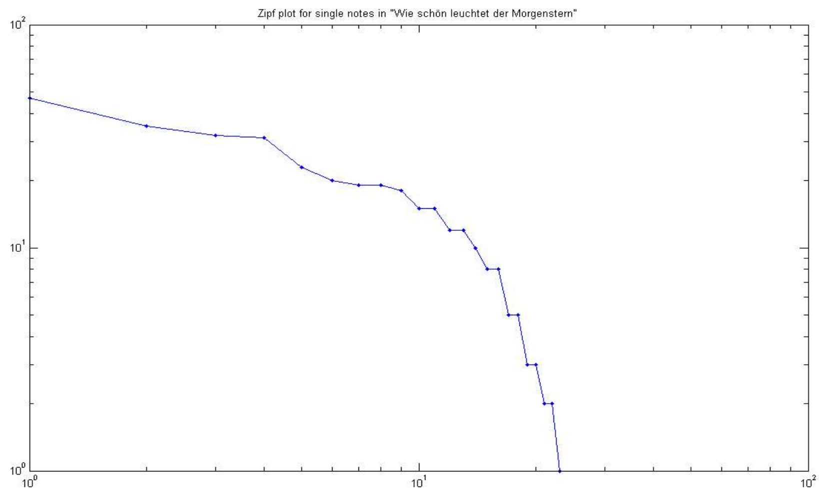


Diagram 5.8. Zipf plot for single notes in "Wie schön leuchtet der Morgenstern" (B378)

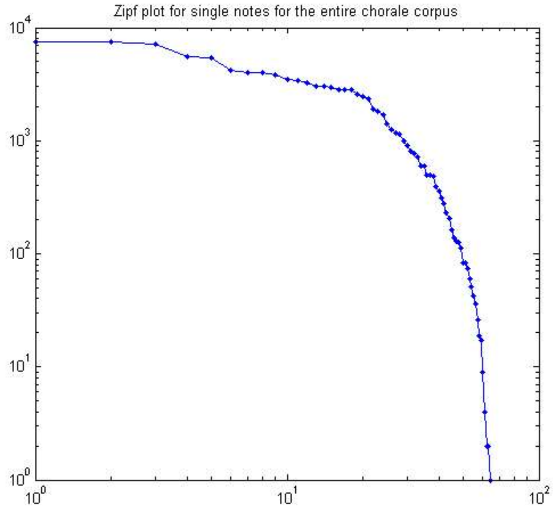


Diagram 5.9. Zipf plot for single notes from the entire chorale corpus

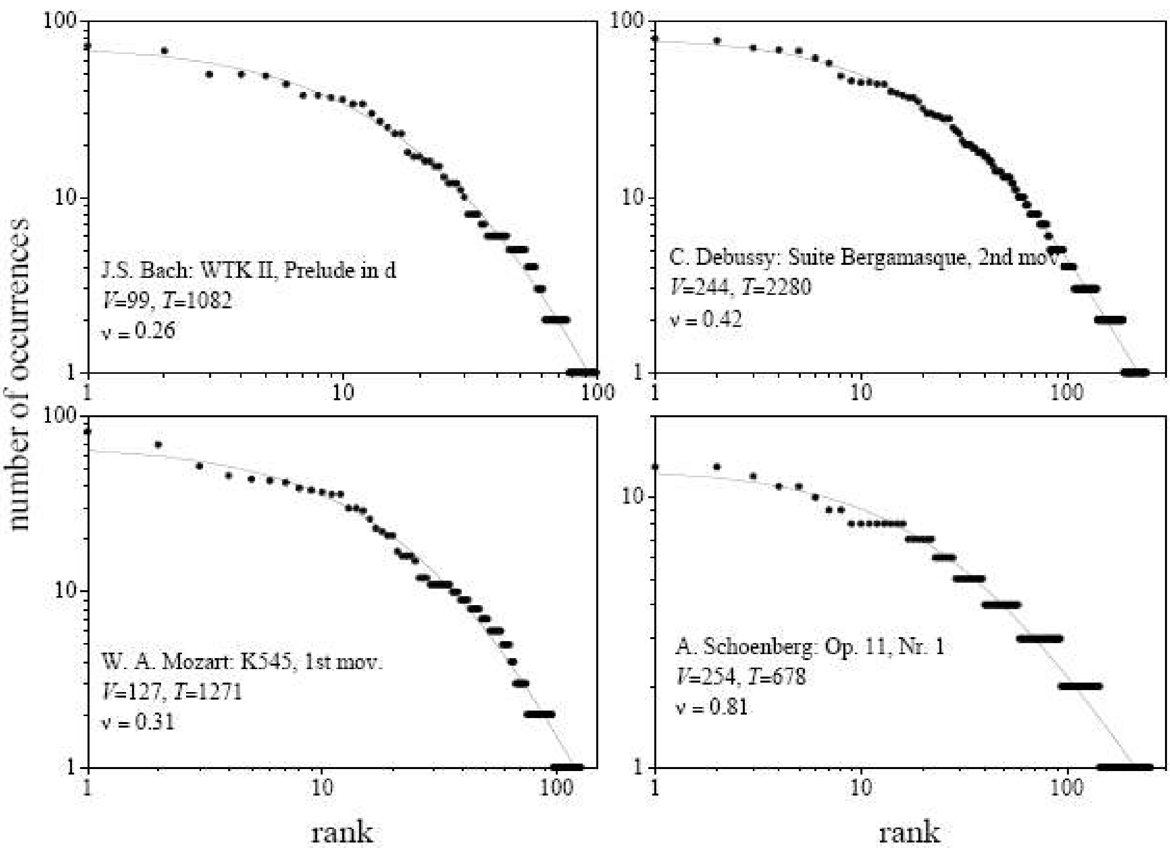


Diagram 5.10, reprinted from Zanette (2005)

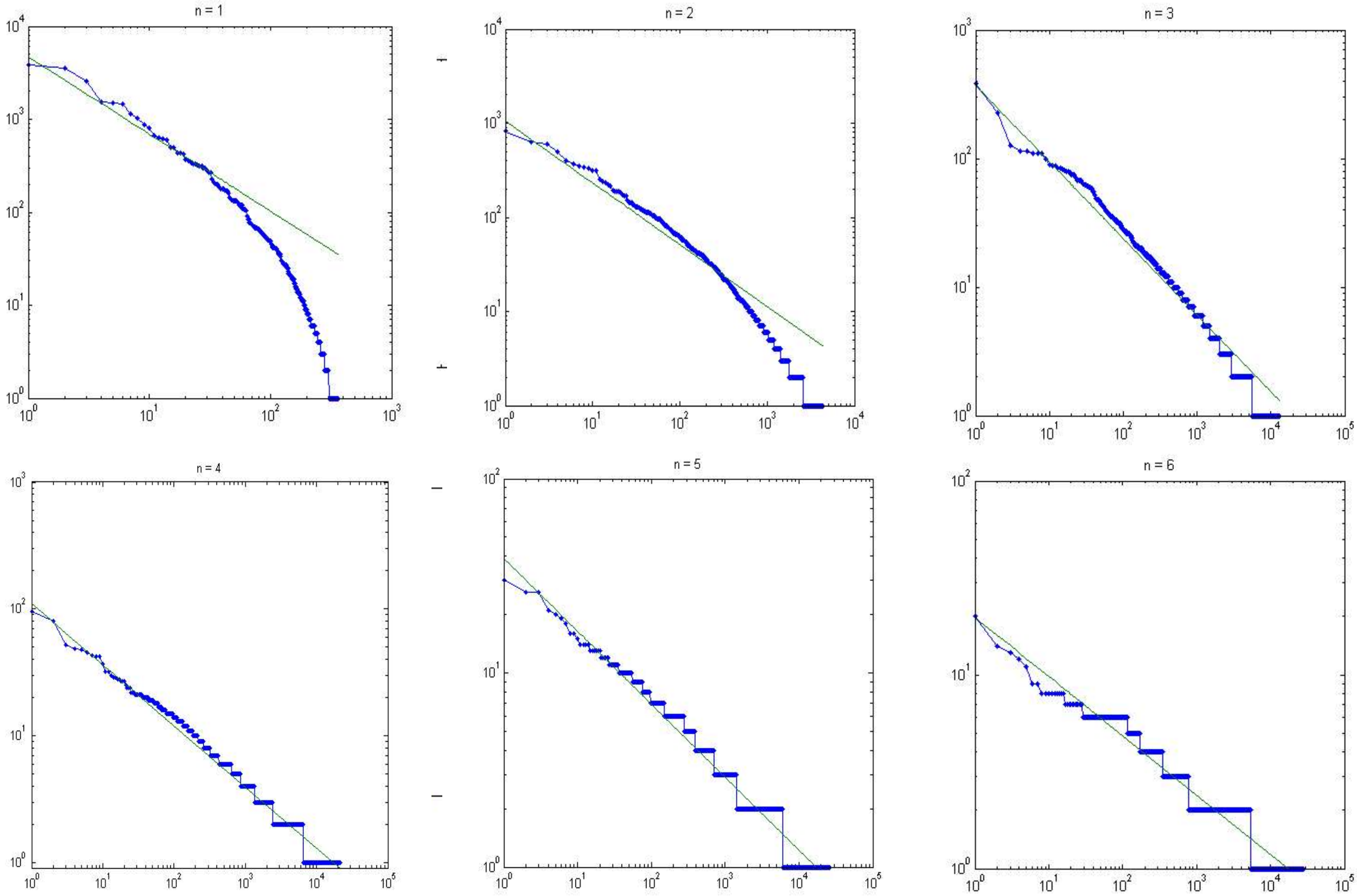


Diagram 5.11. Zipf plots and least square fits with Simon's (1955) formula for n-grams with n = 1,2,3,4,5,6.

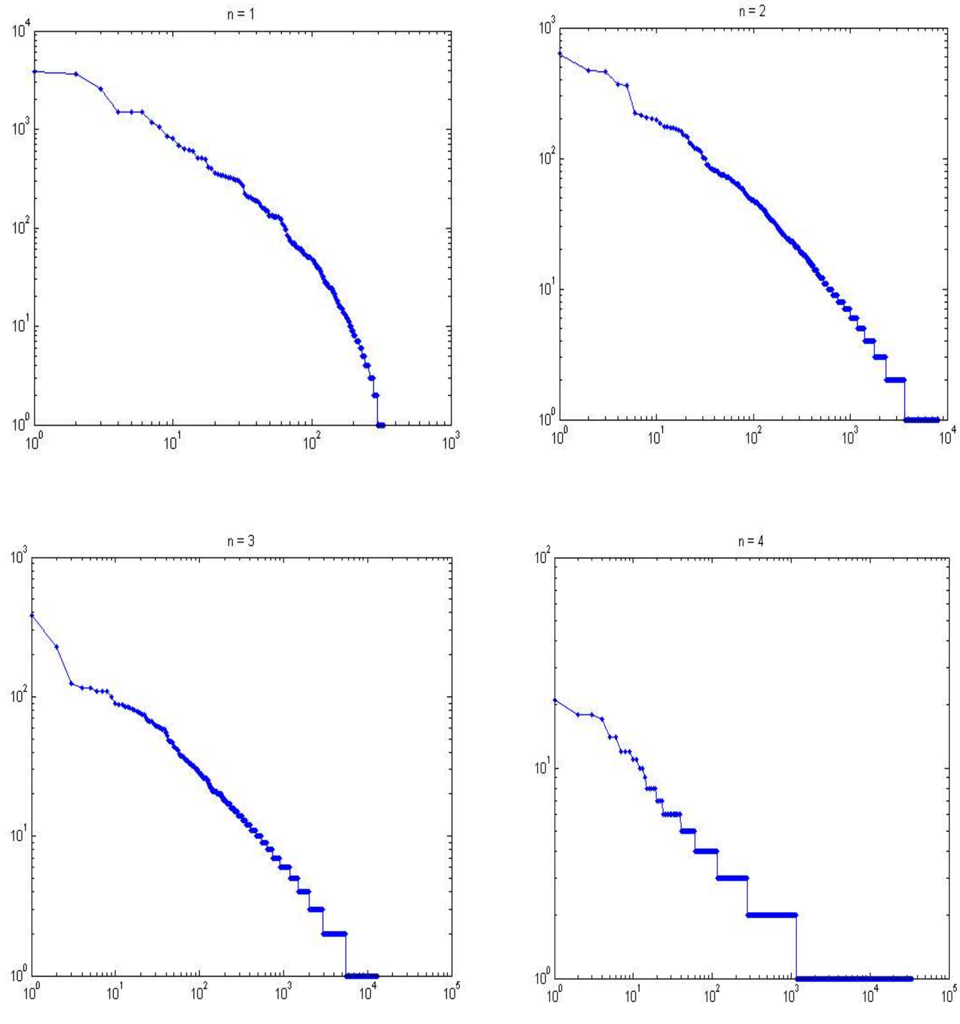
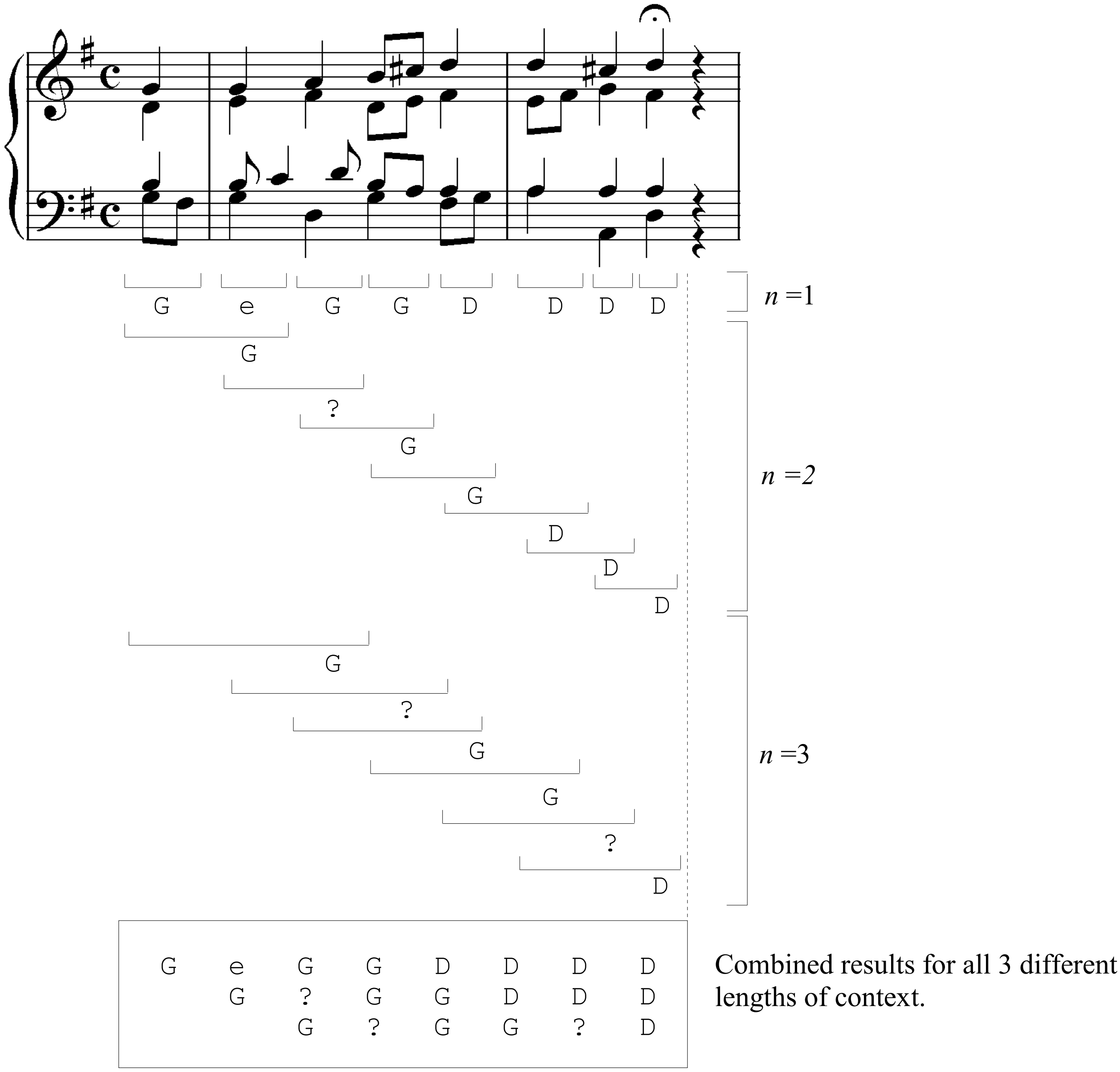


Diagram 5.12. Zipf plots for a random corpus

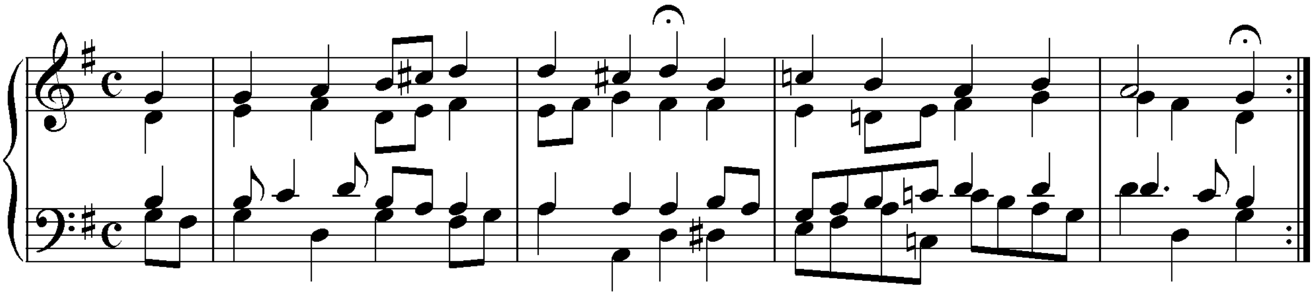
PC set distribution in “Wie schön leuchtet der Morgenstern” (B378)

1	(C.E.G)	29
2	(D.G.B)	14
3	(C.F.A)	10
4	(C.E.A)	8
5	(D.F.B)	5
6	(C.E)	4
7	(C.D.E.G)	4
8	(G.B)	4
9	(E.G.B)	3
10	(C.G)	3
11	(D.F.G.B)	2
12	(D.F.A)	2
13	(C.D.G)	2
14	(C.E.F.A)	2
15	(C.F#.A)	2
16	(D.E.F.B)	2
17	(C.F.A.B)	2
18	(C.D.F.A)	2
19	(D.A)	2
20	(C.F.G.A)	2
21	(C.E.A.B)	2
22	(C.D.F)	1
23	(D.E.G.B)	1
24	(C.D.G.B)	1
25	(C.D.F#.A)	1
26	(E.G)	1
27	(D.G.A.B)	1
28	(G.A)	1
29	(E.G.A.B)	1
30	(C.E.F.G)	1

Table 5.9

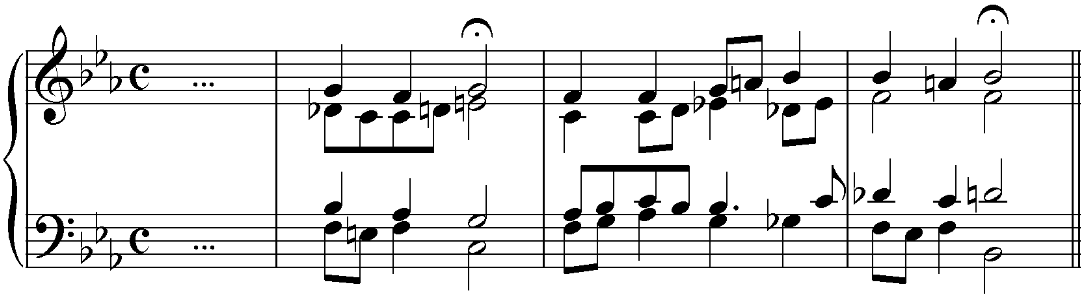


Example 7.2. Combination of key induction results for contexts of the length 1,2,3. The piece is pre-processed with harmonic approximation segmentation.



n=1:	G	e	G	G	D	D	D	D	B	C	G	G	G	G	D	G
n=2:		G	?	G	G	D	D	D	G	e	C	G	G	G	G	G
n=3:			G	?	G	G	?	D	G	G	e	G	G	G	G	G
n=4:				G	G	G	G	?	G	G	G	G	G	G	G	G

Example 7.3. "Ermuntre dich, mein schwacher Geist" (B80)



n=1:	F	f	C	f	f	Eb	F#	bb	F	Bb
n=2:	f	f	f	f	f	?	F#	F#	bb	Bb
n=3:	*f	f	f	f	f	f	*Eb	*bb	bb	bb
n=4:	*Eb	*Eb	f	f	f	f	*Eb	*Eb	*bb	bb

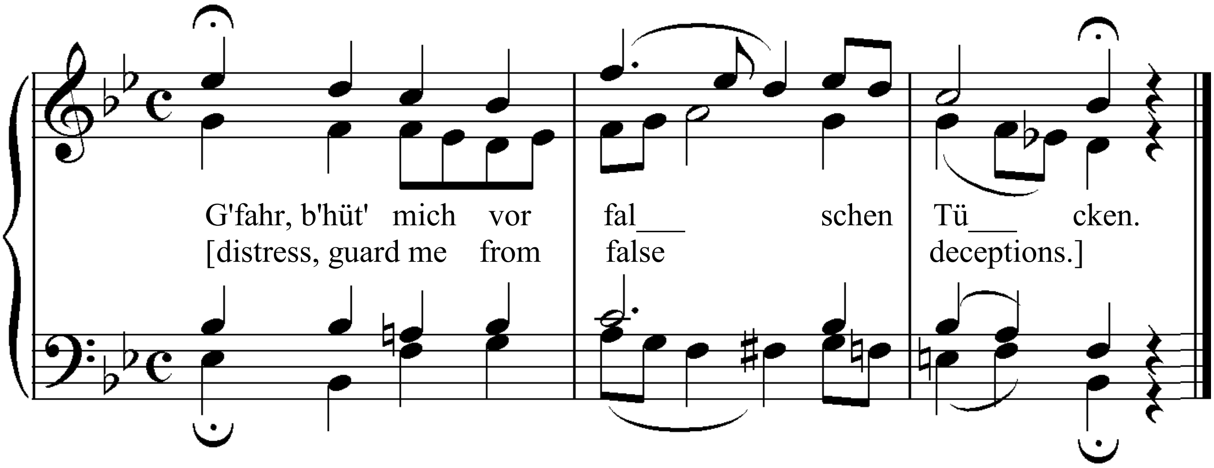
Key expectation

n=2:	F	f	C	f	f	Eb	F#	b	F
n=3:	f	f	f	f	f	Ab	F#	F	bb
n=4:	f	f	f	f	f	f	Eb	bb	bb
	1	2	3	4	5	6	7	8	9

Pc set expectations

	n=1	n=2	n=3
1	(C.F.A)	(C.F.G#)	(C.E.G)
2	(C.F.G#)	(C.E.G)	(C.E.G)
3	(C.E.G)	(C.F.G#)	(C.F.G#)
4	(C.F.G#)	(C.F.G#)	(C.F.G#)
5	(C.F.G#)	(C.E.G)	(C.F.G#)
6	(D#.G.A#)	(C.D#.G#)	(C.D#.G#)
7	(C#.F#.A#)	(C#.F.G#)	(C#.F.A#)
8	(C#.F.A#)	(C#.D#.F#.A#)	(C.F.A)
9	(C.F.A)	(C#.F.A#)	(D.F.A#)

Example 7.4. from "Jesu Leiden, Pein und Tod" (B 194)



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
n=1:	Eb	Bb	F	Bb	g	Eb	F	?	F	Bb	G	Eb	?	F	F	Bb	Bb
n=2:	Eb	Eb	Bb	?	Bb	Bb	Bb	F	F	?	Bb	g	?	?	F	?	Bb
n=3:	Eb	Eb	Bb	Bb	Bb	Bb	Bb	*Bb	F	Bb	Bb	*Bb	g	*f	?	Bb	Bb
n=4:	*Bb	?	Bc	Eb	Bb	Bb	Bb	*Bb	*Bb	Bb	*Bb	*Bb	*Bb	*Bb	*Bb	Bb	Bb

Key expectation

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
n=2:		Eb	Bb	F	Bb	g	Eb	F	F	F	Bb	G	Eb	F	F	F	Bb
n=3:			Eb	Bb	Bb	Bb	Bb	Bb	F	F	Bb	Bb	g	c	F	F	Bb
n=4:				Bb	Bb	Bb	Bb	Bb	B	F	Bb	Bb	Bb	g	f	d	Bb

Pc set expectations

	n=2	n=3	n=4
2	(D#.G.A#)		
3	(D.F.A#)	(D#.G.A#)	
4	(C.F.A)	(D.F.A#)	(D.G.A#)
5	(D.F.A#)	(D.F.A#)	(D.F.A#)
6	(D.G.A#)	(D.F.A#)	(D.F.A#)
7	(D#.G.A#)	(D.F.A#)	(C.F.A)
8	(C.F.A)	(C.D#.F.A)	(D.F.A#)
9	(C.E.G)	(C.E.G)	(C.F.A)
10	(C.F.A)	(C.E.G.A#)	(C.E.F.A)
11	(D.F.A#)	(D.F.A#)	(D.F.A#)
12	(D.G.B)	(D.G.A#)	(D.G.A#)
13	(D#.G.A#)	(D.G.A#)	(D.F.G.A#)
14	(C.E.G)	(C.D#.G#)	(C.D#.G.A)
15	(C.F.A)	(C.F.G.A)	(C.E.G#)
16	(C.F.A)	(C.F.A)	(C.E.F.A)
17	(D.F.A#)	(D.F.A#)	(D.F.A#)

Example 7.5. "In dich hab' ich gehoffet, Herr" (B213) from Bach's St. Matthew Passion

```
n=1:       A       E   b  F# C#        Ab    c#                E    A      E   D    ?
n=2:               A   D b   ?         C#    c#                E    A      A   A    e
n=3:                 *A  D *A         *eb    c#                A    A      A   A  *E
n=4:                   *A *A          *A   *A                  E    A    *A  *A  *A
```

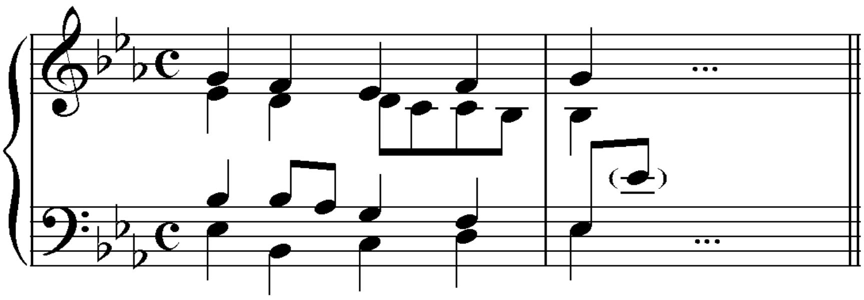
Key expectation

```
n=2:               A   E b   F#        C#    Ab                E    E    A    E    D
n=3:                   A D   b         F#    C#                A    E    A    A    A
n=4:                     A   D         A     eb                E    A    E    A    A

                   1   2 3   4         5     6                      7    8    9    10
```

Pc set expectations

```
1      (C#.E.A)
2      (E.G#.B)      (C#.E.A)
3      (D.F#.B)      (C#.F#.A#)   (C#.F#.A#)
4      (C#.F#.A#)    (D.F#.B)     (D.F#.B)
5      (C#.F.G#)     (C#.F.G#)    (C.D#.G#)
6      (C.D#.G#)     (C#.F.G#)    (D.F.F#.G#)

7      (E.G#.B)      (E.G#.B)     (C#.E.A)
8      (C#.E.A)      (C#.E.A)     (C#.E.G#.A)
9      (E.G#.B)      (C#.E.A)     (C#.E.A)
10     (D.F#.A)      (D.F#.A)     (D.F#.A)
```

Example 7.6. "Es ist genug" (B91)



n=1:	Eb	Bb	Eb	c	c	C	Bb	Eb	Eb
n=2:		Eb	?	?	c	c	Bb	Eb	Eb
n=3:			Eb	Eb	c	Eb	Eb	Eb	Eb
n=4:				?	c	Eb	Eb	Eb	Eb

(here, all “?” symbols characterise ambiguities between c-minor and Eb-major.)

	1	2	3	4	5	6	7	8	9

Expectation:

n=2:	Eb	Bb	Eb	c	c	C	Bb	Eb
n=3:		Eb	Eb	c	c	c	Eb	Eb
n=4:			Eb	c	c	Eb	Eb	Eb

Pc set expectations:

	n=2	n=3	n=4
2	(D#.G.A#)		
3	(D.F.A#)	(D#.G.A#)	
4	(D#.G.A#)	(D#.G.A#)	(D#.G.A#)
5	(C.D#.G)	(C.D#.G)	(C.D#.G)
6	(C.D#.G)	(C.D#.F.G)	(D.G.A#)
7	(C.E.G)	(D.F.B)	(D.F.A#)
8	(D.F.A#)	(D#.G.A#)	(D#.G.A#)
9	(D#.G.A#)	(D#.G.A#)	(D#.G.A#)

Example 7.7. “Jesu Leiden, Pein und Tod” (B194)