Modelling dynamics of key induction in harmony progressions

Martin Rohrmeier
Centre for Music and Science, Faculty of Music, University of Cambridge, UK

Abstract— **This paper presents a statistical model of key induction from given short harmony progressions and its application in modelling dynamics of on-line key induction in harmonic progressions with a sliding window approach. Using a database of Bach's chorales, the model induces keys and key profiles for given harmonic contexts and accounts for related music theoretical concepts of harmonic ambiguity and revision. Some common results from music analytical practice can be accounted for with the model. The sliding window key induction gives evidence of cases in which key is established or modulation is recognised even though neither dominant nor leading tone was involved. This may give rise to a more flexible, probabilistic interpretation of key which would encompass ambiguity and under-determination. In addition, a novel method of harmonically adequate segmentation is presented.**[1]

Keywords— **key induction, harmony, music cognition.**

I. Introduction: Tonality, Cognition and Key Induction

One central aspect in music cognition is key structure [1], [2]. Since key is an abstract structure, the cognition of tonal music involves *key induction* from the musical surface. During the on-line perception of a tonal piece of music a dynamic association of the heard musical elements with respect to an underlying key context is continuously maintained. This association is by no means static nor unambiguous: Like the perception of other structures, such as metre, grouping, or harmonic functions, it changes dynamically throughout the unfolding of the piece and initiates effects of expectation and, consequently, ambiguity and (retrospective) revision. These are central aspects for music perception and musical experiences like the induction of emotional responses [1], [3]. This paper proposes a computational model of key induction, which also aims to account for the dynamic features of on-line key induction, applying a statistical learning approach. In addition, a novel method of segmentation is presented. The research is carried out using a database of Bach's chorales.

Harmonic structures arise from and characterise an underlying key [4]. For instance, the harmony sequence C-a-F-G may induce the key C-major. However, the tonic chord of the key needs not even be present in the harmonic context: e.g. $cm-E_\flat-F$ characterise $B_\flat$-major, $d^{\emptyset 7}$ already characterises c-minor. Hence key induction from given harmonic contexts is not trivial even though it refers to a basic perceptual structure and listening experience. Key induction does not only apply to the present musical context, it may also govern the expectancy of subsequent key and harmonic events. For instance, the C-major example above may raise the expectation of a subsequent C or an a chord, whereas the $d^{\emptyset 7}$ example may lead to expect a G^7 harmony and a subsequent c-minor or possibly a C-major key. Expectancy and key induction are assumed to be mainly based on acquired, schematic musical knowledge, which is assumed to be implicitly learned from multiple exposure to tonal music [2], [5]. Ref. [6], [7] give evidence that a self-organised learning view of tonality is practically plausible using an SOM model. However, a number models of tonality induction only take pitch statistics into account [5], [8], [9] and omit their vertical arrangement which is crucial for harmonic contexts. This model aims to overcome this limitation.

On-line key induction dynamics mainly involve expectations, ambiguity and revision [2]. *Ambiguity* characterises the case in which interpretation is not distinctively possible either because too little information is given or the information equally favours several possibilities: e.g. the chord sequence C-G-C-G might appear as ambiguous between C-major and G-major. However, care has to be given concerning ambiguity: for instance, Agawu [10] gives a nicely imaginary example that, 'strictly' speaking, the beginning of Beethoven's fifth symphony would be highly ambiguous, as the two presented pitch classes G, $E_\flat$ would allow 3 major and 11 minor interpretations. However, this observation holds only from a perspective which involves a 'flat' key-profile. Probabilistic or weighted key profiles like Krumhansl's [5] reduce the number of potential keys drastically. Accordingly, the chord sequence above would score higher for C-major than for G-major in the key induction model presented here.

Revision is a phenomenon which involves the reinterpretation of an already assigned and preferred diachronically maintained analysis due to contradicting evidence. Well known in linguistics (from sentences like "the horse raised past the barn fell" or "the old *man* the boats"), Jackendoff [11] demonstrates the reality and relevance of revision for musical cases, and, similarly, Temperley [2] gives cases for revision concerning the parameters metre, harmony and grouping. In harmony and key induction, revision appears frequently; even modulations are simple, common examples: they frequently involve pivot chords

[1]This research has been carried out under the funding of the German National Academic Foundation and the Microsoft European PhD Scholarship Programme.

which fit both the original and the modulated key and are subject to (functional) reinterpretation. Aldwell & Schachter [4]:ch.32 discuss revision in the context of enharmonic modulation.

II. Model

The model to be presented here suggests a method for key induction from given harmony progressions of a length n. Applying this method of key induction repeatedly to sequent segments of a given musical piece or excerpt realises a sliding window approach to key induction which aims to model the temporal dynamics of on-line key induction outlined above. This sliding window technique may be understood in terms of a very simple model of chord based on-line key induction which respects short-term memory limitations on musical context to be taken into account.

Instead of performing a symbolic functional key analysis of a given short harmony sequence, which could be very complex for certain sequences, a simpler statistical learning approach is proposed. If a ground truth key annotation of whole pieces is given for a large database, all harmony n-grams can be annotated with the key of the piece they appear in. This data yields a key profile for each different n-gram after the n-gram - key pairs have been normalised in order to capture transposed forms of identical harmony sequences. For instance, this accounts for the fact that the progressions a-F-G in C major and $c\sharp$-A-B in E major, are two different realisations of the same normalised progression VI-IV-V for any major key. However, the same progression a-F-G could also appear in a or d minor. Therefore, a key profile for this progression may reflect different values for the keys of C, a, d. Since the same chord progression - key relationship could appear transposed at all 12 keys, a key-invariant relative step-based normalisation of harmony sequences is applied. For this purpose, chord progressions are represented as pc-set sequences [12] which are represented as abstract relative 12-bit vectors independent of any key. These key profiles can be reinterpreted as likelihood/probability profiles for a key context given a pc-set progression. For the computation of the key profiles, the database is split into two subsets, containing all major and all minor chorales, in order to make it possible to compute probability profiles for both modes independently. Given the set S of all chorales, let N_k denote the total number of different n-grams within the major subset if the given k is a major key, or within the minor chorale subset if k is minor. Let $c_s(e)$ be the number of occurrences of an n-gram e in a chorale s, $key(s) \in \mathbb{R}^{24}$ the key of a chorale s which is denoted as the single value 1 in one of the 24 dimensions (major & minor) of the key vector, and $T_a(b)$ the function which transposes a given key vector or n-gram b by an interval a ($a \in \{0, \dots, 11\}$), then the key profile vector of any n-gram e can be expressed as:

$$K(e) = \sum_{i=0}^{11} \sum_{s \in S} \frac{c_s(T_i(e))}{N_{key(s)}} T_{-i}(key(s)) \tag{1}$$

$$\text{with } key(s) \in \{0,1\}^{24}, K(e) \in \mathbb{R}^{24}$$

These key profiles for a given pc-set n-gram can be applied to give a sliding window account of the key dynamics of a given piece. Let e_i^j denote the sequence ($j - i + 1$-gram) of events $e_i \dots e_j$ of pc-sets for some $i \leq j \in \mathbb{N}^+$. Computing all key profiles $K(e_1^n), K(e_2^{n+1}) \dots K(e_{m-n+1}^m)$ over the entire length m of a given piece results in a sliding window account of the key dynamics of the piece. The size of the sliding window can vary between 1 and n.

For event prediction, the probability of an event occurring immediately after a certain context will be described by the *maximum likelihood estimation* [13]:

$$p(e_i^j) = \frac{c(e_i^j)}{\sum_{a \in \xi^{j-i+1}} c(a)} \tag{2}$$

$$p(e_i \mid e_{(i-n)+1}^{i-1}) = \frac{c(e_i \mid e_{(i-n)+1}^{i-1})}{\sum_{a \in \xi} c(a \mid e_{(i-n)+1}^{i-1})} \tag{3}$$

Here, $c(e)$ denotes the total number of occurrences of the n-gram e in S and ξ the set of all different pc-sets in S. Ref. [13] note that this method is simple and may become problematic in contexts of sparse data. However, in this case it will serve for a simple application which does not aim to produce pieces but just analyses the corpus of pieces because unknown contexts cannot occur in this case. In this context, the general 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:

$$p(e_i \mid e_1^{i-1}) \approx p(e_i \mid e_{(i-n)+1}^{i-1}) \tag{4}$$

These can be applied to generate both harmony and key prediction by first generating a pc-set prediction, and then computing the key profile for the n-gram which includes the predicted pc-set.

III. Method

A. Database

The set of Bach's chorales was taken from the online database JSBChorales.net as MIDI-files [14], which contained 521 midi files. Only the 4-part chorale set of the files was taken, and chorales with fewer or more than four part chorales were excluded. Minor errors in two files were corrected; another randomly chosen sample of chorales was manually checked for note errors and did not contain note mistakes. In a few cases, overlapping note durations have been corrected. For the study, only chorales from either Riemenschneider or Kalmus editions were selected. The database (and this subset) contained a number of doublets (including transposition), which have been excluded; for this purpose, a special algorithm was developed in order to handle cases of transposed doublets, using a homomorphism to gain a key independent representation of pc-set transitions, which is described in detail in [15]. In all, the actual chorale corpus contains 386 pieces.

B. Normalisation

Since computing the key profiles for all n-grams requires a ground truth reference key for each chorale, a preprocessing step is necessary to annotate each chorale with its tonal centre. Since there is a large number of modal chorales in the dataset - a fact that is ignored by many computational approaches to Bach's chorales - choices for assignments of tonal centres have to be made for chorales that are not Ionian or Aeolian. It has been considered to be most appropriate to the harmonic structure of the chorales to treat Ionian and Mixolydian cases as major keys, Dorian and Aeolian as minor keys, and Phrygian chorales (25) been assigned to their relative minor key. Arguments for these decisions are given in [15]. The large number of number of modal chorales in the dataset renders standard key finding inapplicable to identify tonal centres.[2] For this purpose, a specialised method, described in detail in [15], assigns tonal centres based on particular music theoretical properties of (potentially modal) final cadences and first chords, which are known to hold reliably for modal music of this time and in particular Bach's chorales [16]. The resulting tonal centre assignments have been tested by expert musicologists and have shown to be correct.

C. Segmentation

It is crucial for the quality of the harmony based key induction model to apply a segmentation algorithm which chooses appropriate harmonies for given segments from pieces from the note-event based (MIDI) database. In general, segmentation is a very hard problem which needs complex algorithms like [17], which even still are not fully applicable to 'real' music. In this case, however some assumptions can be made, based on the fact that the database is a chorale corpus, which reduce segmentation complexity significantly. Bach's chorale compositions mostly consist of a polyphonic composition of four simultaneous voices. Therefore, the problem of inducing adequate harmonies from vertically incomplete sets does virtually not occur (and would average out in the statistical analysis). The problem of assigning appropriate harmonies to each time segment is reduced to choosing the appropriate chord from the set of candidate chords at each time segment.[3] Two different ways of segmentation have been chosen to be applied in the presented model:

a) dense segmentation: If one intends to incorporate all vertical progressions including dissonance treatment into the model, a maximalist *dense segmentation* could be applied. It segments at all time positions where at least one voice/note event changes (comparable to the *full expansion* command in HUMDRUM) (Fig.9.a). 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. This method is appropriate to analyse harmony transitions 'under the microscope', but it can not capture similarities of patterns which just differ slightly, being elaborations of a common underlying harmonic structure.

b) harmony approximation: For a segmentation which represents harmonic progression more abstract on a larger timescale and closer to a cognitive structure, larger segments have to be used and appropriate salient harmony has to be chosen for each segment. Whereas a full, comprehensive reduction to harmonically significant pc-sets is overly complex, it appears appropriate to use metre and consonance as a cognitive cue. Many approaches [18]–[20] sample harmony by selecting only pc-sets at metrically strong positions omitting weaker quavers and semiquavers(*"metrical segmentation"*, Fig. 9.b). This overly simple sampling is severely problematic as it ignores the complexity of the very frequent contrapuntal phenomena in Bach's chorales and samples, e.g., large sets of stressed dissonances or passing notes and not their resolutions.[4]

In order to better approximate to harmonic features of the musical surface, an improved method (*harmony approximation*) employs a selection process which chooses one harmonically representative pc-set from all pc-sets occurring within each crotchet beat.[5] A simple heuristic rule, applied to a given set A of candidate chords within a window of one crotchet, serves the purpose well and handles the great majority of cases adequately.

Rule 1: If the first chord of the set A is dissonant, the least dissonant chord of A will be preferred. If the first chord is consonant or a dominant seventh chord, it will be preferred.

The notion of dissonance is modeled by a heuristic score system for pc-sets. First the pc-set is converted into its (non symmetry-invariant) normal form and its interval vector is computed (after Forte [12].[6] 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 3. Major or minor triads are assigned a value of 2 and major seventh chords (with and without fifth) a value of 1, in order to gain analyses that select preferable harmonically significant chords over 0-scoring incomplete triads. This realises a preference hierarchy of chords which is shown in detail in Fig.10 for all pc-set genera occurring in the chorale corpus.

[2] Retrospectively, Krumhansl's algorithm [5] classified 79.53%, and Temperley's algorithm [2] 20.47% of the chorales correctly.

[3] This assumption holds for a broad range of homophonic and polyphonic music, as long as the composition texture is not largely sparse. The method could be expanded to be applicable to a large set of music by adapting a method to assign harmonies to vertically incomplete pc-sets, for which considerations in [17] might serve as a basis.

[4] For instance,a quaver resolution of G-C-D into G-B-D will be treated as G-C-D.

[5] A 'handmade' rule-based heuristic to correctly identify contrapuntal phenomena like transitory dissonances, neighbour notes, suspensions and their resolutions turned out inapplicable by the practice of Bach's chorales (even though Maxwell [21] appears to have developed an enormously complex predicate-logic for this).

[6] Given the space limits, details cannot be elaborated here, but are described in [15].

Fig. 1.

Fig. 2.

This solution overcomes most problems of a *metrical segmentation*, but few cases which are exceptions of rule 1 produce problematic results: The case of a (as pc-set) 'consonant' suspension is not detected, e.g. a $V^{\frac{6}{4}}$ chord which resolves into a $V^{\frac{5}{3}}$ on the quaver level, cannot be easily and reliably distinguished from an instance of a I-V progression without an algorithmically complex reference to an embedding context. There are also rare vertical passing phenomena which do not have anything to do with any harmonically relevant structure (Fig.1). Fig.2 shows another problematic (though rare) case of passing phenomena where the actual harmony is not even present as one concurrent simple vertical structure. These cases stem from the underlying polyphonic structure and have only few parallels in Bach's chorales, but cannot be treated without a complex rule system. Despite these shortcomings, *'harmony approximation'* proved itself a viable segmentation method to approximate harmony progressions. From a cognitive perspective, the notion of a correlation between stronger metrical position and harmonic relevance is an intuitively acceptable principle [22], [23].

For the computational implementation, the segmentation methods are formalised as follows. Due to the encoding in the MIDI format, a piece is represented as a sequence of note events n_i which are given as vectors of pitch p_i (an integer representing the MIDI pitch), onset o_i and duration time d_i (in beats): $n_i \in Z$ with $Z = \mathbb{R} \times \mathbb{N} \times \mathbb{R}$ and $n_i = \langle o_i, p_i, d_i \rangle$. A piece is characterised as a sequence of note events $\langle n_i \rangle_i \in Z^*$ where Z^* denotes the set of all sequences of members of Z, including the empty sequence ϵ. A segment of the piece is understood as a selection (a sequence) of note events (represented as a set of note event indices), a segmentation as a set 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. ting the MIDI pitch), onset o_i and duration time d_i (in beats): $n_i \in Z$ with $Z = \mathbb{R} \times \mathbb{N} \times \mathbb{R}$ and $n_i = \langle o_i, p_i, d_i \rangle$. A piece is characterised as a sequence of note events $\langle n_i \rangle_i \in Z^*$ where Z^* denotes the set of all sequences of members of Z, including the empty sequence ϵ. A segment of the piece is understood as a selection (a sequence) of note events (represented as a set of note event indices), a segmentation as a set 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 S is a set of segments with an index set M: $S = \{k_i\}_{i \in M}$. This allows for the following characterisation of the method of *dense segmentation*:

$$S_1 = \{k(o_i)\}_{i \in L}, k(t) = \{i \mid o_i \leq t \wedge o_i + d_i > t\} \quad (5)$$

In the case of *metrical segmentation*, only onset times at metrical beat onsets (which are integer values on beat level) are selected:

$$S_2 = \{k(t)\}_{t \in \mathbb{N} \cap [min(o_i); max(o_i)]} \quad (6)$$

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.

$$S_3 = \{h(t)\}_{t \in \mathbb{N} \cap [min(o_i); max(o_i)]} \quad (7)$$

$$\text{with } h(t) = argmax_{k(j), j \in [t; t+1)}(diss(k(j))) \quad (8)$$

For any segmentation S the selected note events are $\{\{n_j\}_{j \in k_i}\}_{i \in M}$. Furthermore, in the segmentation methods described here each segment is represented by a single onset $o(k_i)$ and a single duration $d(k_i) = min\{d_i \mid o_i = o(k_i)\}$. 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:

$$\langle o(k_i), \tau(\{p_j\}_{j \in k_i}), d(k_i) \rangle_{i \in M} \quad (9)$$

$$\text{applying the projection } \tau : \{p_j\}_{j \in k_i} \mapsto \{p_j \text{mod} 12\}_{j \in k_i} \quad (10)$$

IV. Results

A. Sliding Window Tonality Induction

Applying the outlined processing steps of key annotation and segmentation to the pieces in the database, key profiles for the all n-grams in the Bach corpus have been computed. As an example, the sequence $B_\flat - B_\flat^7 - c^{2\cdots1}$ (Fig.8, pc-sets 2-5) yields a key profile vector of $0.387e-3$ for $E_\flat$-major, $0.111e-3$ for $B_\flat$-major, 0.453 for c-minor, 0.340 for g-minor, and 0 otherwise. In relative percentages, this is 0.300 for $E_\flat$-major, 0.086 for $B_\flat$-major, 0.351 for c-minor, 0.263 for g-minor. These results are in accordance with a music theoretical perspective which would favor c-minor and $E_\flat$-major as the most likely keys.

For analyses of the temporal dynamics of key induction, a sliding window approach computes these n-gram key profiles subsequently for each window of the length n of a given piece or excerpt. The resulting temporal key induction changes were visualised into a form of diagram which represents the most likely key for each harmonic segment. In addition, harmonic and key expectations can be computed for each single window using the maximum likelihood estimation (eq.3) described above. These are added to the diagrams below the key induction representation. In the given examples *dense segmentation* and *harmony approximation* were used where indicated.

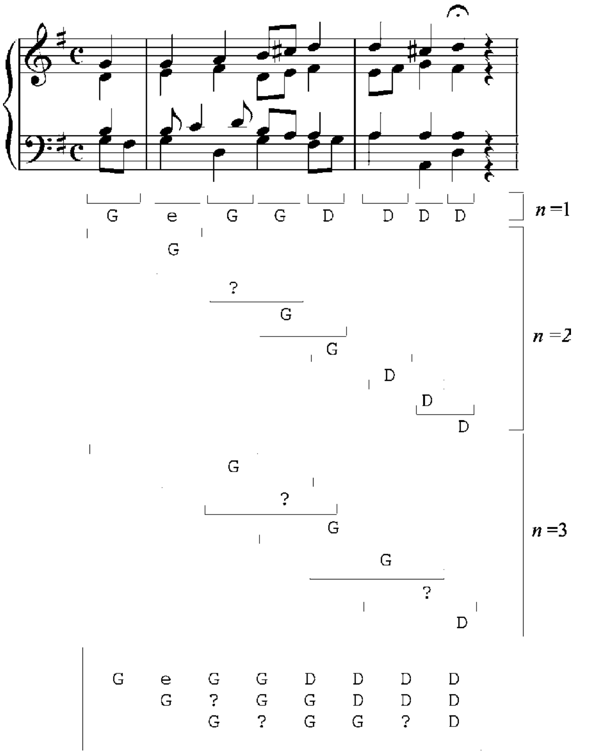

Fig. 3. Single and combined results for 3 different lengths n of context

Figure 3 gives an illustration on how this analysis functions and how the diagrams are constructed. It displays the single segments of the sliding window for context lengths from 1 to 3. In this example *harmony approximation* is applied for segmentation. The key assigned to each segment is denoted by the symbol assigned to the last pc-set; e.g. a symbol in the third row indicates that a context of the previous 2 chords is taken for the key association, a symbol in the first row indicates that only the present chord and no context is taken into consideration. For the cases in which the induced key is ambiguous, the symbol "?" is used in the diagrams. For the detection of these cases a threshold of minimal significant difference has been introduced to mark segments as ambiguities where two or more keys are assigned with fairly equal scores (see [15] for details). In few cases which involve very special/unique harmonic progressions, sparse n-grams occur. In order to avoid problems caused by data sparsity, a threshold has been introduced to identify cases which occur only once and thus might be not representative of the dataset. Key annotations based on sparse data are marked by "$*$" in the diagrams. Further, Fig. 5 gives an example for key and harmony predictions. In the key prediction table, each symbol refers to the predicted key for its segment based on the previous context of the length $n-1$. For instance, a symbol in the row of 3-grams (n=3) denotes that the previous 2 segments create the context that has been applied to yield the key expectation. Accordingly, the table of chord/pc-set expectations display the expected pc-set for each numbered segment given its n-gram context.

Fig. 4. Excerpt from "Ermuntre Dich,mein schwacher Geist"(B80)

Fig. 5. Excerpt from "Jesu, Leiden, Pein und Tod"(B194)

Figure 4 illustrates a practical application of this model and may yield some implications for the role of context in key induction. When the sliding window consists only of one pc-set ($n = 1$) and no context, as one would expect the implied key shifts strongly with nearly every pc-set. Once larger pc-set contexts are taken into account, the estimated key stays more and more stable. With larger n it converges towards the (ground truth) key of the piece due to the way the statistical key annotation is computed. Within larger contextual windows the associated key tends to expose a certain 'inertia': changes which occur on first and second level are carried through much more slowly. Correspondingly, contextual changes tend to be slightly delayed in their effect which may have some cognitive relevance. This can be seen in Fig.5: the sudden modulation (to $b_\flat$-minor) affects the 4th 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.

B. Comparison with music theoretical results

The sliding window approach can be compared to common results in music analytical practice and music theory. Regarding the example at Fig.6, Daniel [16] remarks the double case of an interrupted cadence, which finds an adequate interpretation in the underlying text "falschen Tücken" [false deceptions]. First, a cadential context towards $B\flat$-major is established and then abrogated by


```
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*BbBb *Bb *Bb*Bb *Bb *Bb Bb Bb

       1    2    3  4  5  6    7  8  9 10  11  12  13  14  15 16  17

Key expectation

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

Chord 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#)
```

Fig. 6. Excerpt from "In dich hab' ich gehoffet, Herr" (B213), St. Matthew Passion

D^7 which itself sets up a revised cadential context towards g-minor which is abrogated a second time by $E_\flat$. Then, a continuation to C^7 alludes to F-major before the piece finally reaches the $B_\flat$-major cadence. A comparison of this context with the results of the model shows that this interpretation finds its correlate on a short-term bigram level (induced keys of $B_\flat$ and g (the ambiguous cases marked with "?" are ambiguous between $B_\flat$ and g), then F and finally $B_\flat$) whereas already at the longer contextual levels 3 and 4 the contexts turn out to be weaker.

One compositionally exceptional example (Fig. 7) illustrates prototypical cases of ambiguity, revision and expectation. The first phrase already presents two cases of revision: the established A-major context is reinterpreted to B with the third chord $F\sharp$, which also involves an exceptional accented strong dissonance (which is indeed the only occurrence of this kind throughout the dataset). The next chord $G\sharp^7$, however, effectively destroys the clear cadential context and forces another key revision towards $c\sharp$-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 $D\sharp$ triad. An analysis with the model reproduces some of these effects at bigram level. It is interesting to note here that the harmony progression $E - F\sharp_3$ composed with the strong dissonant 6-5 passing voice in the $F\sharp$ chord is more likely to happen in D major than in b minor. Harmonic expectancies at both cadential segments display the respective expected tonic chords (b at seg.4; D at seg.10).

```
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

Chord 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)
```

Fig. 7. Excerpt from "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

       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

Chord 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#)
```

Fig. 8. Beginning of "Jesu, Leiden, Pein und Tod"(B194)

C. Potential Cognitive Implications

From another perspective, the example in Fig. 8 may rise further cognitive implications on key perception. The computed key profiles display an interesting ambiguity between $E\flat$ major and c minor. Reconsidering the key profile vector example above, it can be seen that, in terms of statistics throughout the body of Bach's chorales, already the transition $E\flat$-$B\flat$-cm suffices to produce a likely interpretation of c-minor. This underlines the tight and ambiguous relationships of minor and relative major keys. But it also may lead to implications about modulation. Though the model had turned out not to modulate too easily on larger contexts, this behaviour suggests an instance of an interesting modulation pattern, as there is no leading tone nor dominant involved. In this case, for instance the pc-set C-D-$E\flat$-G appears to suggest c-minor so distinctively that it overrides the $E\flat$-major context. This result raises some implications with regards to music theory. The classical notion of modulation, e.g. following Schönberg [24], 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. This challenges the "primacy hypothesis" in Brown et al. [25], which postulates key cognition to be governed 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 also encompass ambiguities and under-determination. Similarly to [8], these observations only give rise to hypotheses on temporal characteristics of human key cognition which would have to be investigated experimentally.

V. Conclusions

This paper has presented a statistical model of key induction from harmonic progressions and its application in a sliding window model of the dynamics of on-line key induction and expectation. The analysis has been carried out on the example of a database of Bach's chorales. Furthermore, a novel method of harmonically adequate segmentation suitable for dense polyphonic and homophonic music has been proposed. The model induces appropriate keys and key profiles for given harmonic contexts and it accounts for related music theoretical concepts of ambiguity and revision. Common results from music analytical practice have been reproducible with the model. A sliding window approach to key induction in the chorales revealed that key is established or modulation is recognised even though neither dominant nor leading tone was involved. This challenges Brown's et al. "primacy hypothesis" [25], and gives rise to a more flexible, probabilistic interpretation of key which would encompass ambiguity and under-determination.

VI. Future directions

As outlined above, the model reproduces some central concepts and knowledge in musicology and music theory. From a music psychological perspective it may be a matter of future research to investigate experimentally how well this model compares with actual human *on-line* key perception and with other models. In particular, the results challenging the standard notion of modulation/tonicisation and Brown's "primacy hypothesis" may give rise to an empirical investigation on human perception of modulation. Furthermore, the approach presented does not include on-line built short-term expectancy. It may be a future perspective to incorporate on-line maintained harmonic n-gram transition probabilities for a piece into the model, in a form such as [26], [27] propose. For event prediction, this will have to interact with the long-term prediction computations based on the whole dataset. Possibly, less usual and rare (low-probability and high standard deviation) cases could be treated with a short-term on-line distribution. Besides the presented method could provide a useful tool for music analysis. It could be trained, for instance, on a different or larger corpus (extending the segmentation method to deal with cases of vertically incomplete harmonies, if necessary) or even be applied to investigate stylistic differences in the harmony - induced key relationship between two sufficiently different corpora. Even though Bach's chorales exploit a very large set of possible harmonic progressions, techniques of handling data sparsity as discussed in [13] may be incorporated to gain larger applicability and to generalise and evaluate this model with a larger, more general dataset of tonal music.

References

[1] D. Deutsch, Ed., *The Psychology of Music*, 2nd ed. San Diego: Academic Press, 1999.

[2] D. Temperley, *The cognition of basic musical structures*. Cambridge, MA: MIT Press, 2001.

[3] L. Meyer, *Emotion and Meaning in Music*. London: University of Chicago Press, 1956.

[4] E. Aldwell and C. Schachter, *Harmony and Voice Leading*, 2nd ed. San Diego: Harcourt Brace Jovanovich, 1989.

[5] C. L. Krumhansl, *Cognitive foundations of musical pitch*. New York: Oxford University Press, 1990.

[6] J. J. Bharucha, "Neural nets, temporal composites and tonality," in *The Psychology of Music*, D. Deutsch, Ed. New York: Academic Press, 1998.

[7] B. Tillmann, J. J. Bharucha, and E. Bigand, "Implicit learning of tonality: A self-organizing approach," *Psychological Review*, vol. 107, no. 4, pp. 885–913, 2000.

[8] P. Toiviainen and C. L. Krumhansl, "Measuring and modeling real-time responses to music: the dynamics of tonality induction." *Perception*, vol. 32, no. 6, pp. 741–766, 2003.

[9] D. Temperley, "Bayesian models of musical structure and cognition," *Musicae Scientiae*, vol. 8, pp. 175–205, 2004.

[10] K. Agawu, "Ambiguity in tonal music. a preliminary study," in *Theory, analysis and meaning in music*, A. Pople, Ed., Cambridge, 1994, pp. 86–107.

[11] R. Jackendoff, "Musical parsing and musical affect," *Music Perception*, vol. 9, pp. 199–230, 1991.

[12] A. Forte, *The Structure of Atonal Music*. Yale University Press, 1973.

[13] M. T. Pearce and G. A. Wiggins, "Improved methods for statistical modelling of monophonic music," *Journal of New Music Research*, vol. 33, no. 4, pp. 367–385, 2004.

[14] "Jsbchorales.net," http://www.jsbchorales.net/sets.html, April 14 2005, the history of corrections can be found at http://www.jsbchorales.net/correc.html.

[15] M. Rohrmeier, "Towards modelling movement in music: Analysing properties and dynamic aspects of pc set sequences in bach's chorales," Master's thesis, University of Cambridge, 2005.

[16] T. Daniel, *Der Choralsatz bei Bach und seinen Zeitgenossen. Eine historische Satzlehre*. Köln: Dohr, 2000.

[17] B. Pardo and W. P. Birmingham, "Algorithms for chordal analysis," *Computer Music Journal*, vol. 26, no. 2, pp. 27–49, 2002.

[18] H. Hild, J. Feulner, and D. Menzel, "Harmonet: a neural net for harmonising chorales in the style of j. s. bach," in *Advances in Neural Information Processing 4*, R. P. Lippmann, J. E. Moody, and D. S. Touretzky, Eds. San Francisco: Morgan Kaufmann, 1992, pp. 267–274.

[19] M. E. Jø rgensen and S. T. Madsen, "Harmonisation of bach chorales. kbs project report." Daimi. University of Aarhus., Tech. Rep., 2002.

[20] D. Ponsford, G. Wiggins, and C. Mellish, "Statistical learning of harmonic movement," *Journal of New Music Research*, vol. 28, pp. 150–177, 1999.

[21] J. H. Maxwell, "An expert system for harmonizing analysis of tonal music," in *Understanding Music With AI:Perspectives on Music Cognition*, M. Balaban, K. Ebcioglu, and O. Laske, Eds. Cambridge, Massachusetts: MIT Press, 1992, p. 335353.

[22] F. Lerdahl and R. Jackendoff, *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press, 1983.

[23] D. Temperley and D. Sleator, "Modeling meter and harmony: A preference rule approach," *Computer Music Journal*, vol. 15, no. 1, pp. 10–27, 1999.

[24] A. Schönberg, *Structural Functions of Harmony*. London: Faber and Faber, 1969.

[25] H. Brown, D. Butler, and M. Jones, "Musical and temporal influences on key discovery," *Music Perception*, vol. 11, pp. 371–407, 1994.

[26] D. Conklin, "Prediction and entropy of music," Master's thesis, Department of Computer Science, University of Calgary (Technical Report 199039014), 1990.

[27] M. T. Pearce and G. A. Wiggins, "Expectation in melody: The influence of context and learning," *Music Perception*, vol. 23, no. 5, pp. 377–405, 2006. [Online]. Available: http://caliber.ucpress.net/doi/abs/10.1525/mp.2006.23.5.377

APPENDIX

Fig. 9. Different methods of segmentation: (a) dense segmentation (b) metric segmentation (c) harmonic approximation

pc set	score
$(C.E.G)$	2
$(C.D_\sharp.G)$	2
$(C.D_\sharp.F_\sharp.G_\sharp)$	1
$(C.D.F_\sharp)$	1
$(C.F)$	0
$(C.E)$	0
$(C.D_\sharp)$	0
(C)	0
$(C.D_\sharp.F.G_\sharp)$	-1
$(C.D.G)$	-1
$(C.D_\sharp.F_\sharp)$	-1
$(C.F_\sharp)$	-1
$(C.D_\sharp.F)$	-1
$(C.D.F)$	-1
$(C.D)$	-1
$(C.D_\sharp.F_\sharp.A)$	-2
$(C.D.F.G_\sharp)$	-2
$(C.D_\sharp.F.G)$	-2
$(C.D.F.G)$	-2
$(C.D.E.G)$	-2
$(C.E.F_\sharp)$	-2
$(C.D.E)$	-2
$(C.D.E.G_\sharp)$	-3
$(C.E.G_\sharp)$	-3
$(C.D.F_\sharp.G_\sharp)$	-4
$(C.C_\sharp.F.G_\sharp)$	-4
$(C.D_\sharp.E.G_\sharp)$	-4
$(C.C_\sharp.E.G_\sharp)$	-4
$(C.D_\sharp.E.G)$	-4
$(C.D.E.F_\sharp)$	-4
$(C.E.F)$	-4
$(C.C_\sharp.F)$	-4
$(C.D_\sharp.E)$	-4
$(C.C_\sharp.E)$	-4
$(C.C_\sharp)$	-4
$(C.D_\sharp.F_\sharp.G)$	-5
$(C.E.F.G)$	-5
$(C.C_\sharp.E.G)$	-5
$(C.D.D_\sharp.G)$	-5
$(C.F.F_\sharp)$	-5
$(C.C_\sharp.F_\sharp)$	-5
$(C.D.D_\sharp)$	-5
$(C.C_\sharp.D_\sharp)$	-5
$(C.E.F_\sharp.G)$	-6
$(C.D.F_\sharp.G)$	-6
$(C.C_\sharp.F.G)$	-6
$(C.C_\sharp.D_\sharp.G)$	-6
$(C.D_\sharp.F.F_\sharp)$	-6
$(C.D_\sharp.E.F_\sharp)$	-6
$(C.D.D_\sharp.F_\sharp)$	-6
$(C.C_\sharp.D_\sharp.F_\sharp)$	-6
$(C.D.E.F)$	-6
$(C.D.D_\sharp.F)$	-6
$(C.C_\sharp.D_\sharp.F)$	-6
$(C.C_\sharp.E.F)$	-8
$(C.C_\sharp.F.F_\sharp)$	-9
$(C.C_\sharp.D_\sharp.E)$	-9

Fig. 10. Dissonance ratings for all pc-set genera occurring in Bach's chorales