STATISTICAL PARSING FOR HARMONIC ANALYSIS OF JAZZ CHORD SEQUENCES

Mark Granroth-Wilding, Mark Steedman
School of Informatics
University of Edinburgh, Edinburgh, UK
mark.granroth-wilding@ed.ac.uk, steedman@inf.ed.ac.uk

ABSTRACT

Analysing music resembles natural language parsing in requiring the derivation of structure from an unstructured and highly ambiguous sequence of elements, whether they are notes or words. Such analysis is fundamental to many music processing tasks, such as key identification and score transcription.

The focus of the present paper is on harmonic analysis. We use the three-dimensional tonal harmonic space developed by [4, 13, 14] to define a theory of tonal harmonic progression, which plays a role analogous to semantics in language. Our parser applies techniques from natural language processing (NLP) to the problem of analysing harmonic progression. It uses a formal grammar of jazz chord sequences of a kind that is widely used for NLP, together with the statistically based modelling techniques standardly used in wide-coverage parsing, to map music onto underlying harmonic progressions in the tonal space.

Using supervised learning over a small corpus of jazz chord sequences annotated with harmonic analyses, we show that grammar-based musical parsing using simple statistical parsing models is more accurate than a baseline Markovian model trained on the same corpus.

1. INTRODUCTION

Musical meter, melody and harmonic progressions exhibit hierarchical structure, similar to the structure found in the prosody and syntax of language. In linguistics, this is analysed using tree diagrams to represent recursive divisions of constituents in a passage of text or speech down to the level of individual words. Research in natural language processing has developed an armoury of techniques to process this structure, many of which may be equally applied to interpretation of music.

Analysing a sentence's syntactic structure, or *parsing* the sentence, is often a prerequisite to semantic interpretation. The analysis is typically highly ambiguous, even for moderately long sentences. The field of statistical parsing aims to overcome the ambiguity by reference to knowledge of commonly occurring constructions. In music, a similar sort of structural analysis over a sequence of notes is fundamental to tasks such as key identification and can play an important role in others, like song recognition. These tasks in general depend on both harmonic and metrical analyses.

We focus here on harmonic analysis. We use a three-dimensional tonal harmonic space ([4, 13, 14, 15]). This representation provides the basis for a theory of tonal harmonic progression. The framework allows us to analyse the relationships between the chords underlying a passage of music and the relationship of the notes to their underlying chords. We treat the analysis of the tonal relations between chords analogously to the logical semantics of a sentence. By defining a representation of movements in the tonal space in a form similar to logical representations of natural language semantics, we are able to apply techniques from NLP directly to the problem of harmonic analysis.

We use a formal grammar of jazz chord sequences in a formalism based closely on one used for NLP. We then use modelling techniques commonly applied to the task of statistical parsing of natural language sentences with such grammars to map music, in the form of chord sequences, onto its underlying harmonic progressions in the tonal space. In the present paper, we omit the details of the representation of tonal space movements as formalized in the grammar (the semantics), and the syntactic component of the grammar. Instead we introduce the structures that the grammar is designed to analyse and focus on the statistical parsing techniques and their performance on the harmonic analysis task.

We use supervised learning over a small corpus of chord sequences (76 songs, ∼3k chords) of jazz standards from lead sheets used by performers, annotated by hand with harmonic analyses that we treat as a gold standard. We describe some experiments comparing the use of grammar-based musical parsing aided by simple statistical parsing models from NLP to a baseline Markovian model that also produces an analysis in the tonal space. We show that the grammar-based model performs better than the baseline model at producing a tonal space analysis matching the hand-annotated gold standard.

[18] presented a small, context-free syntactic grammar of jazz chord sequences designed to capture twelve-bar blues chord sequences. [19] further developed the blues grammar, using a syntactic formalism and language of harmonic analysis that form the basis for those used in the present work to construct a wider-coverage grammar

of tonal jazz chord sequences. The present paper uses statistical models to apply the grammar to an analysis task. [5] and [16] have proposed a syntactic model of harmonic structure closely related to that we employ. The experiments we present provide further support for structured approaches to musical analysis and the use of techniques adapted from NLP.

2. MUSICAL SYNTAX

The syntax of tonal harmony and that of natural language can both be analysed using tree structures, and both have been claimed to feature formally unbounded embedding of structural elements ([10, 12, 18, 16]).

2.1. Cadences

The *cadence*, built from tension-resolution relationships between chords, forms the basic unit of harmonic structure. Large structures, which we will refer to as *extended cadences*, are made up of successive tension-resolution patterns chained together. There are two main varieties of cadence. An *authentic* (or *perfect*) cadence consists of a tension chord rooted a perfect fifth above its subsequent resolution. The tension chord is called a *dominant* chord. A *plagal* cadence consists of a tension chord rooted a perfect fourth above its resolution. Such a tension chord is a *subdominant* chord. In both cases, the resolution chord is classified as a *tonic* chord.

The identification of an occurrence of a chord with its role in one of these structures is referred to as its *function*. It partly establishes the chord's place in the harmonic structure of the musical passage. A particular chord type, say a G major triad, may function either as a dominant or subdominant tension chord, or as a tonic resolution, on different occurrences within the same piece.

A tension chord may resolve by the expected interval to another chord which is also cadential and thus creates a further tension and itself resolves subsequently. Such a definition is recursive, and extended cadences can accordingly be indefinitely extended. This kind of extension may be applied to either type of cadence, though it is uncommon with the plagal cadence.

An example is shown in the form of a tree in figure 1. A cadence Dm^7 G^7 C has two possible interpretations: it may contain a recursive dominant relation, as in the figure, or be a substitute transcription of the perfect cadence F^6 G^7 C. When, as in this case, the recursion reaches back further, however, only the former interpretation explains for the relation between the seemingly tonally distant tension chords and their eventual resolution (here the A^7 and its resolution to C).

In some cases, a tension chord may not immediately reach the resolution it calls for. The *unresolved* dominant cadence Dm^7 G^7, for example, creates an expectation of a tonic C chord. It may be interrupted by a further cadence, A^7 Dm^7 G^7, creating the same expectation, whereupon *both* cadential expectations/tensions will be resolved by the *same* tonic C, as in

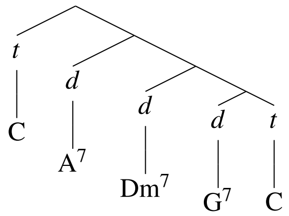

Figure 1. An extended authentic cadence, a typical example of (tail) recursion in music. The A^7 acts as a dominant resolving to the Dm^7, which in turn resolves by the same relation to G^7, which then resolves to the tonic C.

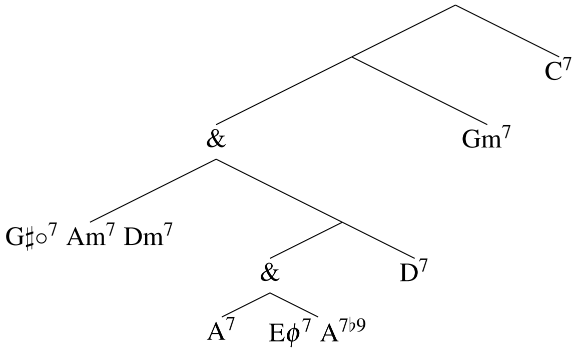

Figure 2. Tree representing the embedded structure of unresolved cadences in *Call Me Irresponsible*, coordination of constituents marked by &. The Dm^7 chord is left unresolved until the Gm^7 and the A^7 until the D^7. The entire example is in fact further embedded in the song: the eventual resolution to the tonic F is not reached until after another cadence of similar structure.

C (Dm^7 G^7) (A^7 Dm^7 G^7) C

We term this operation *coordination* by virtue of its similarity to right-node raising coordination in natural language. For example, in *Keats bought and will eat beets*, *beets* satisfies the expectations of both *bought* and *will eat*.

Coordinated cadences may themselves be embedded in coordinated cadences, as in the example taken from *Call Me Irresponsible* shown in figure 2. Once again, a similar form of embedding occurs in natural language examples like *Keats (certainly eats) but ((may or may not) cook) beets*.

Dominant function chords are often partially, though never unambiguously, distinguished by the addition of notes outside the basic triad. In particular, the *dominant seventh*, realized by the note two semitones below the octave, enhances the cadential function of a dominant chord and heightens expectation of the resolution. However, a dominant may omit this note and the same note (or rather, one indistinguishable from it in equal temperament) may even appear in chords not functioning as dominants.

2.2. The Jazz Sublanguage

The typical size and complexity of the cadence structures discussed above varies with musical period and genre. Tonal jazz standards are of particular interest for this form of analysis for several reasons.

First, they tend to feature large extended cadences, often with complex embedding. Second, they contain many well-known *contrafacts*, harmonic variations of a familiar piece, created using a well-established system of harmonic substitutions, embellishments and simplifications.

Finally, jazz standards are rarely transcribed as full scores, but are more analytically notated as a melody with accompanying chord sequence. Analysing the harmonic structures underlying chord sequences, rather than streams of notes, avoids some difficult practical issues such as voice leading and performance styles, but still permits discovery of the kind of higher-level structures we are concerned with. They therefore provide a convenient starting point for our investigation.

Our study focusses on the analysis of harmonic structure in chord sequences of jazz standards. This is not to say that the approach to analysis is not applicable beyond this domain or even that it depends on analysing chord sequences. Our grammar's lexicon, introduced in brief in section 4, is specific to the genre, though a lexicon suitable for another tonal harmonic genre would have much in common.

3. A MODEL OF TONALITY

In analysing the roles of pitch in music, it is important to distinguish between *consonance*, the sweetness or harshness of the sound that results from playing two or more notes at the same time, and harmony, which is the dimension relevant to the phenomenon that we have already alluded to as tension (and the creation of expectation) and resolution (or its satisfaction). Both of these relations over pitches are determined by small whole-number ratios, and are often confounded. However, they arise in quite different ways.

3.1. Consonance

The modern understanding of consonance originates with Helmholtz ([6]), who explained the phenomenon in terms of the coincidence and proximity of the secondary overtones and difference tones that arise when simultaneously sounded notes excite real non-linear physical resonators, including the human ear itself, inducing harmonics or secondary tones. To the extent that an interval's most powerful secondary tones exactly coincide, it is perceived as consonant or sweet-sounding. To the extent that any of its secondaries are separated in frequency by a small enough difference to *beat* at a certain rate, it is perceived as dissonant, or harsh.

Thus, for the diatonic semitone only very high-frequency, low-energy overtones coincide, so it is weakly consonant, while the two fundamentals themselves produce beats in the usual musical ranges, so it is strongly dissonant. For the perfect fifth, with a frequency ratio of 3/2, all its most powerful secondaries coincide, so it is strongly consonant.

E^-	B^-	$F\sharp^-$	$C\sharp$	$G\sharp$	$D\sharp$	$A\sharp$	$E\sharp^+$	$B\sharp^+$
C^-	G^-	D^-	A	E	B	$F\sharp$	$C\sharp^+$	$G\sharp^+$
$A\flat^-$	$E\flat^-$	$B\flat^-$	F	C	G	D	A^+	E^+
$F\flat^-$	$C\flat^-$	$G\flat^-$	$D\flat$	$A\flat$	$E\flat$	$B\flat$	F^+	C^+
$D\flat\flat^-$	$A\flat\flat^-$	$E\flat\flat^-$	$B\flat\flat$	$F\flat$	$C\flat$	$G\flat$	$D\flat^+$	$A\flat^+$

Figure 3. Part of the space of note-names (adapted from [13, 14])

This theory successfully explains the experience of consonance and dissonance in chords, and the effects of chord inversion. We ignore the issue of consonance, unlike [9, 11], and are interested instead in the somewhat orthogonal issue of harmony.

3.2. Harmony

The tonal harmonic system also derives from combinations of small integer pitch ratios. However, the harmonic relation is based solely on the first three prime ratios in the harmonic series: ratios of 2, 3 and 5 (the octave, perfect fifth and major third). The tuning based on these intervals is known as *just intonation*.

3.2.1. Just Intonation

In just intonation, an interval can be represented as a frequency ratio defined as the product $2^x \cdot 3^y \cdot 5^z$, where x, y, z are positive or negative integers. It has been observed since [4] that the harmonic relation can therefore be visualized as an infinitely extending discrete three-dimensional space with these three prime factors as generators. Since notes separated by octaves are essentially equivalent for tonal purposes, it is convenient to project the space onto the $3, 5$ plane. We adopt this theory in the form in which it was formally developed by Longuet-Higgins ([13, 14]), shown in figure 3 in its two-dimensional projection.

[15] observed that all diatonic scales are convex sets of positions, and defined a Manhattan distance metric over this space. According to this metric, it will be observed that the major and minor triads, such as CEG and CE♭G, when plotted in this space are two of the closest possible clusters of three notes. The triad with added major seventh is the single tightest cluster of four notes. The triads and the major seventh chord are stable unambiguous chords that raise no strong expectations and are of the kind that typically end a piece. Chords like the diminished chord and the dominant seventh are more spread out, a difference vital to the induction of harmonic expectation and its satisfaction.

3.2.2. Equal Temperament

Over several centuries, an approximation of the tonal harmonic space was gradually adopted, first by slightly mistuning the fifths to equate all the positions with the same label in figure 3, and then by further distorting the major thirds, to equate enharmonic equivalents (C with B♯, D♭♭, etc.). The 12 tones of the diatonic octave are spaced evenly, so that all the semitones are (mis)tuned to the same ratio of $\sqrt[12]{2}$.

This system of *equal temperament* has the advantage that all keys and modes can be played on the same instrument without retuning. In the tonal space, the result is a distortion of the pitches so that the infinite space is projected onto a finite space of just 12 points, looping in both dimensions. Each point is (potentially infinitely) tonally ambiguous as to which point in the infinite justly intoned space of figure 3 it denotes. Thus, equal temperament makes the interpretation of tonal relations ambiguous. Its advantage, however, is that it allows the hearer to *resolve* this tonal ambiguity.

It is important to realize that ambiguous equally tempered music is unconsciously interpreted in terms of the full tonal space of harmonic distinctions, just as a theoretically infinitely ambiguous two-dimensional photograph is interpreted as a three-dimensional scene. We perform this disambiguation explicitly in our analyses by mapping equal-temperament chord sequences onto paths through the justly intoned tonal space.

3.3. Domain for Analysis

In our grammar for jazz chord sequences, we take the full two-dimensional tonal space as the semantic domain of harmonic analysis. A harmonic interpretation of a piece is the path through the tonal space traced by the roots of the chords.

If we establish that there is a dominant-tonic relationship between two chords, we know that the underlying interval between the roots is a perfect fifth, a single step to the left in the space. Likewise, a subdominant-tonic relationship dictates a perfect fourth, a rightward step. Where no tension-resolution relationship exists, as between a tonic and the first chord of a cadence that follows it, we assume a movement to the most closely tonally related instance of the chord root.

Figure 4 shows an example of a harmonic interpretation of an extended cadence as a tonal space path. The perfect fifth relationship between each dominant seventh chord and its resolution is reflected in the path. There is no tension-resolution relationship between the first two chords (a tonic and the start of the cadence), so the path proceeds to the closest instance of the A. Consequently, the path ends at a different C to the origin, not distinguished by equal temperament from the starting point.

By identifying the syntactic structure of the harmony, that is the recursive structure of tension-resolution relationships between pairs of chords, we can produce the path through the space that it dictates for the chord roots of the progression.

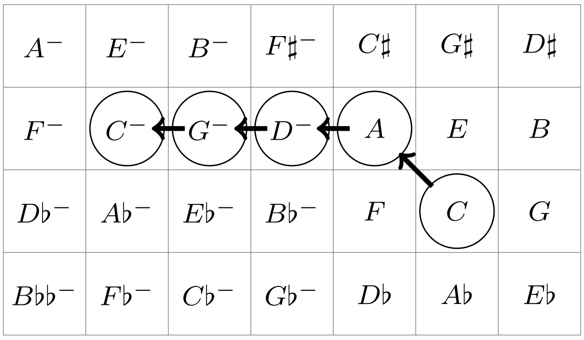
A^- E^- B^- $F\sharp^-$ $C\sharp$ $G\sharp$ $D\sharp$
F^- C^- G^- D^- A E B
$D\flat^-$ $A\flat^-$ $E\flat^-$ $B\flat^-$ F C G
$B\flat\flat^-$ $F\flat^-$ $C\flat^-$ $G\flat^-$ $D\flat$ $A\flat$ $E\flat$

Figure 4. A tonal space path for the extended cadence: C A^7 D^7 Gm^7 C.

4. A GRAMMAR FOR JAZZ

Combinatory Categorial Grammar (CCG) is a lexicalized grammar formalism. A CCG lexicon contains *categories* that are assigned to the words of a sentence which specify constraints on the structures in which the word's semantics may combine with that of surrounding words. Once a category has been chosen for each word, a small set of *combinators* may be used to produce a semantics for the whole sentence from that of the individual words. An adaptation of CCG to the parsing of harmony was introduced by [19]. We use here a further development of that formalism and introduce a statistical parsing model and an implementation that were missing there.

We have hand-crafted a lexicon containing categories suitable for assigning harmonic interpretations to chords. Our musical CCG grammar contains several combinators, similar to those used for parsing natural language. Each item in the lexicon is a schema that generalizes over chord roots. When it is assigned to a chord, it assumes the chord's root and thenceforth applies its constraints relative to that root.

For example, a schema *Dom* is used to interpret a chord as having a dominant function, including recursive dominant sevenths, as described above. It constrains its subsequent resolution to be rooted a perfect fifth below it. Its semantics represents a leftward step in the tonal space.

Another schema *Ton* interprets a chord as a tonic chord and may serve as the resolution to a *Dom* category. Further categories are included to handle subdominant chords, substitutions (such as the tritone substitution), passing chords, and so on.

Figure 5 shows a full CCG derivation of the cadence from *Call Me Irresponsible*, the structure of whose harmonic semantics was shown in figure 2, this time with a final tonic resolution appended. (In fact, this resolution is not reached until after another, similar cadence structure.) We do not describe the grammar's lexicon and combina-

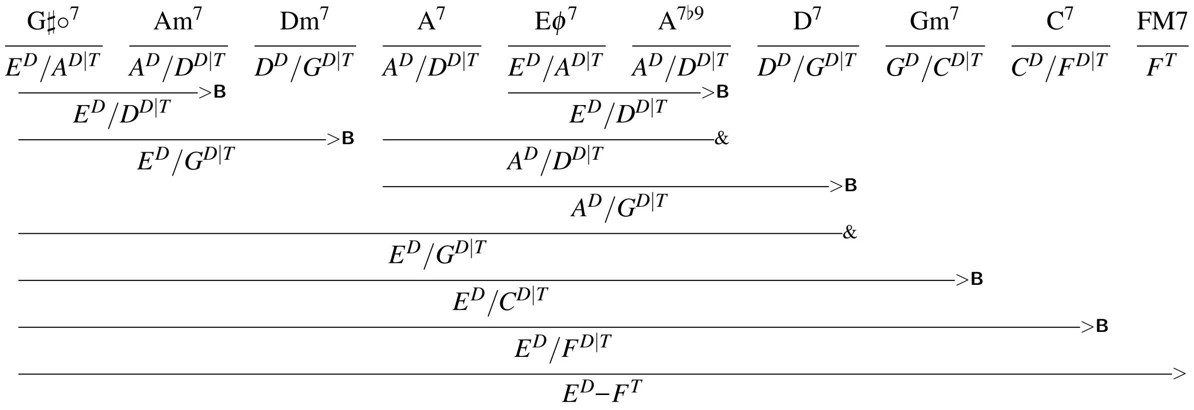
G♯o7 Am7 Dm7 A7 Eϕ7 A7♭9 D7 Gm7 C7 FM7
E^D/A^D|T A^D/D^D|T D^D/G^D|T A^D/D^D|T E^D/A^D|T A^D/D^D|T D^D/G^D|T G^D/C^D|T C^D/F^D|T F^T
>B
E^D/D^D|T
>B
E^D/G^D|T
>B
E^D/D^D|T
&
A^D/D^D|T
>B
A^D/G^D|T
&
E^D/G^D|T
>B
E^D/C^D|T
>B
E^D/F^D|T
>
E^D–F^T

Figure 5. CCG derivation of part of *Call Me Irresponsible*. The top line is the input chord sequence. Beneath, a single category is assigned to each chord. Subsequent lines combine categories as licensed by the grammar's combinators. Not shown here, each category has an associated logical form representing tonal space points or movements.

tors any further here, but include this derivation merely to give a flavour of how an interpretation is produced from lexical categories.

The lexicon is deliberately specific to the genre we wish to analyse. Another lexicon could be constructed with which to interpret the harmonic relations of another tonal harmonic genre and would have a number of categories in common. A lexicon for European baroque music, for example, would not use all of the substitution categories included in the jazz lexicon and would require some additional categories to reflect different conventional expressions of the perfect cadence.

5. STATISTICAL PARSING MODELS

Just as with natural language parsing, the lexical ambiguity of interpretation of chord sequences prohibits exhaustive parsing to deliver every syntactically well-formed interpretation of a sequence. Moreover, we require some means by which to identify the most plausible interpretations among a huge number of interpretations permitted by the grammar.

It is usual in parsing natural language to use statistical models based on a corpus of hand-annotated sentences to rank the permissible interpretations of the input and of parts of it. Such techniques can be used to reduce the search space during parsing and speed up parsing by ignoring seemingly improbable interpretations early in the process. [2, 8, 20] have applied statistical parsing techniques from NLP to chord sequence parsing and other tasks for folksong domains. This paper shows that such methods can be extended to the present class of musical grammars.

5.1. Jazz Corpus

We acquired the statistics used by our models from a small corpus of jazz chord sequences. We chose the sequences from available lead sheets, excluding certain sequences that could not be analysed using our grammar, due to limitations of the lexicon (e.g. rare substitutions not covered).

We annotated the chord sequences by hand, assigning to every chord a category from the lexicon of the jazz grammar. Since CCG is a lexicalized grammar formalism, the assignment of categories to chords contains a large amount of information constraining the parse. We also added annotations of the points where coordination occurs, providing sufficient information to define a unique tonal space analysis of every sequence.

The corpus consists of 76 annotated sequences, totalling roughly 3000 chords. It contains no held-out test set: all models are tested using cross-validation (see section 6.2). We plan to make the corpus publicly available in the future.

5.2. Adaptive Supertagging

Supertagging is a technique, related to part of speech (POS) tagging, useful as a first step in parsing with lexicalized grammars like CCG ([17]). Probabilistic sequence models, using statistics about short windows of sequences, are employed to choose CCG categories from the lexicon for each word. In music, as in natural language, the choice of a category representing a plausible interpretation of a chord depends on the analysis of potentially distant parts of the sequence (long-distance dependencies). In practice, short-distance statistics can often reliably rule out at least the most improbable interpretations.

A bad choice of categories could make it impossible to parse the sequence. The *adaptive supertagging* algorithm ([3]) allows categories considered less probable by the supertagger to be used in such cases. First, the supertagger assigns to each word (or chord) a small set of what its model dictates are the most probable categories and the parser attempts to find a full parse with these categories. If it fails, the supertagger supplies some more, slightly less probable categories and the parser tries again. This is repeated until the parser succeeds or we give up (for example, after a set number of iterations). If multiple full parses are found in one iteration, the single most probable one is chosen.

Many types of probabilistic sequence model can be used as a supertagging model. We use a hidden Markov model (HMM) with states representing categories. The state emissions of the model are not the chords themselves, but a pair of the chord type and the interval between this chord's and the previous chord's roots. This has the effect of making the model account only for relative pitch. We trained the model by maximum likelihood estimation over the annotated categories from the corpus described above.

We performed some initial experiments with higher-order Markov models (n-gram models) which suggested that they do not perform any better than the HMM we use here when trained on this small corpus. We expect that the model would benefit from the use of higher-order statistics given a larger training set.

5.3. Parsing Models

[7] adapted the generative probabilistic parsing models of probabilistic context-free grammars (PCFG) to CCG. Using a corpus of gold-standard parsed sentences, probabilities are estimated for expansions at internal nodes in the derivation tree. These probabilities are used to estimate a probability for every subtree produced during the derivation.

In our experiments, we use a model like that of [7] to parse chord sequences, which we refer to as PCCG. During parsing, the model is used to assign a probability to internal nodes in the derivation: that is, every combination of categories by a combinator. A *beam* is applied to internal nodes: all but the most probable derivations, according to the parsing model's probabilities, are removed.

A second model uses the supertagger with the adaptive supertagging algorithm described above to narrow down the choice of lexical categories available to the parser. The parser then proceeds just as in PCCG. We call this model St+PCCG.

Using both models, we allow the parser a fixed amount of time to parse a particular sequence before giving up, chosen such that most parses complete within the time.

5.4. Baseline Model

In an attempt to quantify the contribution made by restricting interpretations to those that are syntactically well formed under the jazz grammar, we have constructed a model which assigns tonal space interpretations without using the grammar. We use an HMM very similar to that described above as a supertagger model, which directly assigns a tonal space point to each chord, instead of assigning categories to chords and parsing to derive a tonal space path. The representation of the chord sequence is identical to the supertagger's.

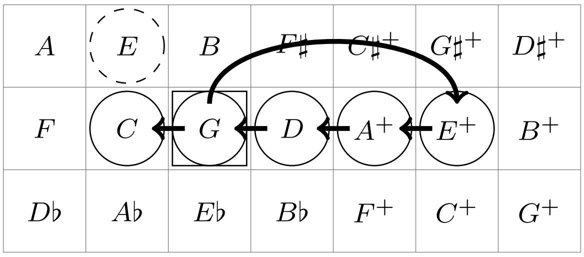

Figure 6. Tonal space analysis for the coordinated cadence G^7 E^7 A^7 Dm^7 G^7 C. The initial G^7 (square) is followed by a jump not to the closest point that equal temperament maps to E (dashed), but a more distant one. This must be so because the resolution of this G^7 and that at the end of the second cadence are constrained to be the same.

We can define a naive, deterministic procedure to construct a tonal space interpretation for a chord sequence as follows: for each chord, choose from the infinite set of points mapped by equal temperament to the chord's root the point that is closest to the previous point on the path. The states of the model are constructed to represent deviations from this naive path.

There are two reasons why such deviations from the naive path are required for valid analyses. First, there may be a substitution (like the tritone substitution), so that the surface chord's root is not the root of the chord in the analysis. Second, the correct disambiguation of the equal-temperament note may not be the point closest to the previous point, as happens at points of coordination (as in the example in figure 6).

The naive procedure identifies the correct tonal space point in most cases and deviations are usually small. The HMM's state labels are of the form $(x_{sub}, y_{sub}, x_{block}, y_{block})$. The pair (x_{sub}, y_{sub}) identifies the relationship between the equal-temperament projection of the chord root in the analysis and that notated, thus modelling chord substitution. $(0,0)$ is most common; the tritone substitution would result in $(2,1)$. (x_{block}, y_{block}) accounts for cases where the tonal relation between this and the last root is not closest, measured by Manhattan distance. It represents the distance from this initial estimate to the root in terms of a number of horizontal and vertical cycles of the equal temperament 4×3 space. The states of the HMM only include those tonal relations observed in the training data.

The model is trained in the same way as the supertagger, only this time the training data is chord sequences paired with their annotated tonal space paths. We refer to this model as HmmPath.

Unlike the supertagger, this model's results are not filtered by the parser for grammaticality. PCCG and St+PCCG will completely fail to assign a path in cases where a full parse cannot be found. This may be because the beam removes all derivations that permit a grammatical interpretation of the full sequence, or, in the case of St+PCCG, because the supertagger fails to suggest a set of lexical categories from which a full interpretation can be derived. HmmPath will assign some path to any sequence, since it is not limited to returning grammatical interpretations.

5.5. Adaptive Supertagging with Backoff

PCCG and ST+PCCG will both fail to produce an interpretation for some sequences. This means that, however high quality the returned paths are, the overall score is inevitably pulled down by the failure to interpret the chords of the omitted sequences.

A fourth model combines the coverage of HMMPATH with the precision of the grammatical models in an aggressive form of *backoff*. First, if a result can be obtained from ST+PCCG it is used. Otherwise, HMMPATH is applied instead. We refer to this combined model as ST+PCCG+HMMPATH.

6. EXPERIMENTS

6.1. Evaluation

We evaluate all models on the basis of the tonal space path they produce with highest probability. For evaluation, paths are transformed from a list of tonal space coordinates to a list of vectors between adjacent points. This has the effect that if a path makes an incorrect jump, it is penalized only for that mistake and not for all subsequent points. Each point also has an associated chord function, which is included in the evaluation.

We align the list of vectors optimally with that of the gold-standard tonal space path from the annotated corpus using the Levenshtein algorithm, with points where the vector is correct but the function wrong, or vice versa, incurring a cost of 0.5.

We report precision, recall and f-score of the aligned paths. Precision is defined as the proportion of points returned by the model that correctly align with the gold standard. Recall is the proportion of points in the gold standard that are correctly retrieved by the model. Again, we give a score of 0.5 to points with either the vector or function correct but not both. F-score is the harmonic mean of precision and recall.

$$P = Aligned/(Aligned + Inserted)$$

$$R = Aligned/(Aligned + Deleted)$$

$$F = 2PR/(P+R)$$

6.2. Model Comparison

All models were trained on the jazz corpus described above, containing 76 fully annotated sequences. Since we cannot afford to hold out a test set, we used 10-fold cross-validation. Each experiment was run 10 times, with $\frac{9}{10}$ of the data used to train the model and the remaining $\frac{1}{10}$ used to evaluate that model. Thus, all data is used for evaluation, but no model is tested on data that it was trained on. We report the results combined from all partitions.

The evaluation of the tonal space path is performed in every case only on the path returned by the model with highest probability.

7. RESULTS

The results of the four experiments are reported in table 1.

Although PCCG has the full set of lexical categories available to it, its results are all lower than ST+PCCG. This is because we needed to apply a more aggressive beam during parsing in order to handle the wider choice of interpretations at the lexical level. It seems, then, that the supertagger is a necessity for practical parsing and is doing a good job of cutting down the parser's search space.

ST+PCCG produces high-precision results, because, unlike HMMPATH, it can only produce results that are permitted by the grammar and fails when it can find no such result. As we would expect, the addition of the backoff reduces the model's precision, but improves its recall. Since ST+PCCG rarely fails to produce a result on this dataset, the backoff has little impact on the overall result. However, ST+PCCG+HMMPATH is robust in that it is guaranteed always to produce some result.

As described in section 5.1, we included in our corpus only sequences to which it is possible to assign a valid harmonic interpretation using our grammar. The results we report here for the models that use the grammar are therefore higher than we would expect if applying the technique to chord sequences sighted in the wild. In this case, we would expect the benefit of the backoff to become clearer, since the PCCG model would more often fail to find an analysis.

We draw two key conclusions from the results. First, they show that HMMPATH is a reasonable model to back off to when no grammatical result can be found. Second, they show that the use of a grammar to constrain the paths predicted by an HMM supertagger substantially improves over the purely short-distance information captured by a pure HMM-based model.

8. CONCLUSION

We have described a parser that uses a formal grammar of a kind employed in NLP, and statistically based modelling techniques of a kind standardly used in wide-coverage natural language parsers, to map music onto their harmonic interpretation, represented as harmonic progressions in the two-dimensional tonal space. The jazz harmony corpus we used to train our models is small, but experience with CCG parsing for NLP shows that these techniques will scale to larger datasets ([1, 3]).

The parsing model is built using supervised learning over a small corpus of jazz chord sequences hand-annotated with harmonic analyses. We found that our grammar-based musical parser, using a simple statistical parsing model, more accurately reproduced the gold-standard interpretations than a baseline Markovian model.

Model	*Precision (%)*	*Recall (%)*	*F-score (%)*	*Coverage (%)*
HMMPATH	81.1	87.7	84.3	100
PCCG	78.1	83.1	80.6	94.7
ST+PCCG	**86.0**	90.8	88.3	98.7
ST+PCCG+HMMPATH	85.7	**92.0**	**88.7**	100

Table 1. Evaluation of each model's prediction of tonal space paths using 10-fold cross-validation on the jazz corpus.

This may be taken as further evidence suggesting that music and language have a common origin in a uniquely human system of interpersonal communication.

We have described models to analyse sequences of chords expressed in a textual form. A certain amount of analysis has already gone into the process of producing these chord symbols: a human has divided the music into time segments of constant harmony, selected the most prominent notes, and narrowed down the range of possible chord roots somewhat. We intend to continue this work by constructing a model that incorporates these tasks into the analysis process, accepting note-level input (in MIDI form, for example) and suggesting possible interpretations in the way the supertagger component of our parsing model does.

9. REFERENCES

[1] M. Auli and A. Lopez, "Training a log-linear parser with loss functions via softmax-margin," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Edinburgh: ACL, 2011, pp. 333–343.

[2] R. Bod, "Memory-based models of melodic analysis: Challenging the gestalt principles," *Journal of New Music Research*, vol. 31, pp. 27–37, 2002.

[3] S. Clark and J. R. Curran, "Wide-coverage efficient statistical parsing with CCG and log-linear models," *Computational Linguistics*, vol. 33, pp. 493–552, 2007.

[4] L. Euler, *Tentamen novae theoriae musicae ex certissismis harmoniae principiis dilucide expositae*. Saint Petersberg Academy, 1739, tonnetz p.147.

[5] W. B. Haas, M. Rohrmeier, R. C. Veltkamp, and F. Wiering, "Modeling harmonic similarity using a generative grammar of tonal harmony," in *Proceedings of the Tenth International Conference on Music Information Retrieval (ISMIR)*. International Society for Music Information Retrieval (ISMIR), 2009, pp. 1–6.

[6] H. Helmholtz, *Die Lehre von den Tonempfindungen*. Braunschweig: Vieweg, 1862, trans. Alexander Ellis (1875, with added notes and appendices) as *On the Sensations of Tone*.

[7] J. Hockenmaier and M. Steedman, "Generative models for statistical parsing with Combinatory Categorial Grammar," in *Proceedings of the 40th Meeting of the ACL*, Philadelphia, PA, 2002, pp. 335–342.

[8] A. Honingh and R. Bod, "Convexity and well-formedness of musical objects," *Journal of New Music Research*, vol. 34, pp. 293–303, 2005.

[9] P. Johnson-Laird, O. Kang, and Y. C. Leong, "On musical dissonance," *Music Perception*, vol. 29, 2011, in press.

[10] A. Keiler, "Two views of musical semiotics," in *The Sign in Music and Literature*, W. Steiner, Ed. Austin TX: University of Texas Press, 1981, pp. 138–168.

[11] C. Krumhansl, *Cognitive Foundations of Musical Pitch*. Oxford: Oxford University Press, 1990.

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

[13] C. Longuet-Higgins, "Letter to a musical friend," *The Music Review*, vol. 23, pp. 244–248, 1962.

[14] ——, "Second letter to a musical friend," *The Music Review*, vol. 23, pp. 271–280, 1962.

[15] C. Longuet-Higgins and M. Steedman, "On interpreting Bach," *Machine Intelligence*, vol. 6, pp. 221–241, 1971.

[16] M. Rohrmeier, "Towards a generative syntax of tonal harmony," *Journal of Mathematics and Music*, vol. 5, pp. 35–53, 2011.

[17] B. Srinivas and A. Joshi, "Disambiguation of super parts of speech (or supertags): Almost parsing," in *Proceedings of the International Conference on Computational Linguistics*. Kyoto: ACL, 1994.

[18] M. Steedman, "A generative grammar for jazz chord sequences," *Music Perception*, vol. 2, pp. 52–77, 1984.

[19] ——, "The blues and the abstract truth: Music and mental models," in *Mental Models in Cognitive Science*, J. Oakhill and A. Garnham, Eds. Erlbaum, 1996, pp. 305–318.

[20] D. Temperley, *Music and Probability*. Cambridge, MA: MIT Press, 2007.