A Probabilistic Model for Chord Progressions

Jean-François Paiement
IDIAP Research Institute
Rue du Simplon 4
Case Postale 592
CH-1920 Martigny
Switzerland
paiement@idiap.ch

Douglas Eck
Dep. of Computer Science
and Operations Research
University of Montreal
CP 6128 succ Centre-Ville
Montréal, QC
H3C 3J7, Canada
eckdoug@iro.umontreal.ca

Samy Bengio
IDIAP Research Institute
Rue du Simplon 4
Case Postale 592
CH-1920 Martigny
Switzerland
bengio@idiap.ch

ABSTRACT

Chord progressions are the building blocks from which tonal music is constructed. Inferring chord progressions is thus an essential step towards modeling long term dependencies in music. In this paper, a distributed representation for chords is designed such that Euclidean distances roughly correspond to psychoacoustic dissimilarities. Estimated probabilities of chord substitutions are derived from this representation and are used to introduce smoothing in graphical models observing chord progressions. Parameters in the graphical models are learnt with the EM algorithm and the classical Junction Tree algorithm is used for inference. Various model architectures are compared in terms of conditional out-of-sample likelihood. Both perceptual and statistical evidence show that binary trees related to meter are well suited to capture chord dependencies.

1 Introduction

Probabilistic models for analysis and generation of polyphonic music would be useful in a broad range of applications, from contextual music generation to on-line music recommendation and retrieval. However, modeling music in general involves long term dependencies in time series that have proved very difficult to capture with traditional statistical methods. Note that the problem of long-term dependencies is not limited to music, nor to one particular probabilistic model (Bengio et al., 1994). This difficulty motivates our exploration of chord progressions. Chord progressions constitute a fixed, non-dynamic structure in time and thus can be used to aid in describing long-term musical structure.

One of the main features of tonal music is its organization around *chord progressions*. A chord is a group of three or more notes (generally five or less). A chord progression is simply a sequence of chords. In general, the chord progression itself is not played directly in a given musical composition. Instead, notes comprising the current chord act as central polarities for the choice of notes at a given moment in a musical piece. Given that a particular temporal region in a musical piece is associated with a certain chord, notes comprising that chord or sharing some harmonics with notes of that chord are more likely to be present. In typical tonal music, most chord progressions are repeated in a cyclic fashion as the piece unfolds, with each chord having in general a length equal to integer multiples of the shortest chord length.

Chord changes tend to align with metrical boundaries in a piece of music. Meter is the sense of strong and weak beats that arises from the interaction among a hierarchy of nested periodicities. Such a hierarchy is implied in Western music notation, where different levels are indicated by kinds of notes (whole notes, half notes, quarter notes, etc.) and where bars establish measures of an equal number of beats Handel (1993). For instance, most contemporary pop songs are built on four-beat meters. In such songs, chord changes tend to occur on the first beat, with the first and third beats (or second and fourth beats in syncopated music) being emphasized rhythmically. Chord progressions strongly influence melodic structure in a way correlated with meter. For example, in jazz improvisation notes perceptually closer to the chord progression are more likely to be played on metrically-accented beats with more "dissonant" notes played on weaker beats. For a complete treatment of the role of meter in musical structure see Cooper and Meyer (1960).

This strong link between chord structure and overall musical structure motivates our attempt to model chord sequencing directly. The space of sensible chord progressions is much more constrained than the space of sensible melodies, suggesting that a low-capacity model of chord progressions could form an important part of a system that analyzes or generates melodies. As an example, consider blues music. Most blues compositions are variations of a basic *same* 12 bar chord progression[1]. Identification of that chord progression in a sequence would greatly contribute to genre recognition.

In this paper we present a graphical model that captures the chord structures in a given musical style using as

[1]In this paper, chord progression are considered relative to the key of each song. Thus, transposition of a whole piece has no effect on our analysis.

evidence a limited amount of symbolic MIDI[2] data. One advantage of graphical models is their flexibility, suggesting that our model could be used either as an analytical or generative tool to model chord progressions. Moreover, a model like ours can be integrated into a more complex probabilistic transcription model (Cemgil, 2004), genre classifier, or automatic composition (Eck and Schmidhuber, 2002).

Cemgil (2004) uses a somewhat complex graphical model that generates a mapping from audio to a pianoroll using a simple model for representing note transitions based on Markovian assumptions. This model takes as input audio data, without any form of preprocessing. While being very costly, this approach has the advantage of being completely data-dependent. However, strong Markovian assumptions are necessary in order to model the temporal dependencies between notes. Hence, a proper chord transition model could be appended to such a transcription model in order to improve polyphonic transcription performance.

Raphael and Stoddard (2003) use graphical models for labeling MIDI data with traditional Western chord symbols. In this work, a Markovian assumption is made such that each chord symbol depends only on the preceding one. This assumption seems sufficient to infer chord symbols, but we show in Section 4 that longer term dependencies are necessary to model chord progressions by themselves in a generative context, without regard to any form of analysis.

Lavrenko and Pickens (2003) propose a generative model of polyphonic music that employs Markov random fields. Though the model is not restricted to chord progressions, the dependencies it considers are much shorter than in the present work. Also, octave information is discarded, making the model unsuitable for modeling realistic chord voicings. For instance, low notes tend to have more salience in chords than high notes (Levine, 1990).

Allan and Williams (2004) designed a harmonization model for Bach chorales using Hidden Markov Models. A harmonization is a particular choice of notes given a sequence of chord labels. While generating excellent musical results, this model has to be provided sequences of chords as input, restricting its applicability in more general settings. Our work goes a step further by modeling directly chord progressions in an unsupervised manner. This allows our proposed model to be directly appended to any supervised model without the need for additional data labeling.

In Section 2, we introduce a similarity measure for chords guided by psychoacoustic considerations. A probabilistic model for chord progressions is then described in Section 3. The model uses our proposed similarity measure for chords to distribute the probability mass of the training set to unseen events appropriately. In Section 4.1 we evaluate the likelihood of the model against reference data. Finally, in Section 4.2 we show that chord sequences generated by the proposed model are more realistic than the ones generated by simpler models in terms of global dependencies.

[2] In our work, we only consider notes onsets and offsets in the MIDI signal.

2 Chord Similarities

The generalization performance of a generative model depends strongly on how observed data is represented. If we had an infinite amount of data, we could simply represent each chord as a distinct observation without including in the model any specific knowledge about psychoacoustic similarities between chords. Unfortunately, typical symbolic music databases are very small compared to the complexity of the polyphonic music signal. Hence, a useful statistical model for chord progressions has to include notions of psychoacoustic similarity between chords in order to redistribute efficiently a certain amount of probability mass to unseen events during training.

One possibility we chose not to consider was to represent directly some attributes of Western chord notation such as “minor”, “major”, “diminished”, etc. Though inferring these chord qualities could have aided in building a similarity measure between chords, we found it more convenient to build a more general representation directly tied to the acoustic properties of chords. Another possibility for describing chord similarities is set-class theory, a method that has been compared to perceived closeness (Kuusi, 2001) with some success.

In this paper, we consider a simpler approach where each group of observed notes forming a chord is seen as a single timbre (Vassilakis, 1999). From this timbre information, we derive a continuous distributed representation where perceptually similar chords tend also to be close in Euclidean distance. One can design a graphical model that directly observes these continuous representations of chords. However, this approach would suffer from two major drawbacks. First, it is unnatural to compress discrete information in a continuous space; one could easily think of a one-dimensional continuous representation that would overfit any discrete dataset. Second, since the set of likely chords is finite, one wants to observe directly discrete variables with a finite number of possible states.

Our proposed solution to this problem is to convert the Euclidean distances between chord representations into probabilities of substitution between chords. Chords can then be represented as individual discrete events. These probabilities can be included directly into a graphical model for chord progressions, as described in Section 3. It is interesting to note that the problem of considering similarities between discrete objects in statistical models is not restricted to music and encompasses a large span of applications, including natural language processing and biology.

2.1 Continuous Representation

The frequency content of an idealized musical note i is composed of a fundamental frequency $f_{0,i}$ and integer multiples of that frequency. The amplitude of the h-th harmonic $f_{h,i} = hf_{1,i}$ of note i can be modeled with geometric decaying ρ^h, with $0 < \rho < 1$ (Valimaki et al., 1996).

Consider the function

$$m(f) = 12(\log_2(f) - \log_2(8.1758))$$

that maps frequency f to MIDI note $m(f)$. Let $\mathbb{X} =$

$\{\mathcal{X}_1 \ldots \mathcal{X}_s\}$ be the set of the s chords present in a given corpus of chord progressions. Then, for a given chord $\mathcal{X}_j = \{i_1, \ldots, i_{t_j}\}$ with t_j the number of notes in chord $\mathcal{X}_j$, we associate to each MIDI note n a perceived loudness

$$l_j(n) = \max_{h \in \mathbb{N}, i \in \mathcal{X}_j} (\{\rho^h | \mathtt{round}(m(f_{h,i})) = n\} \cup \{0\}) \tag{1}$$

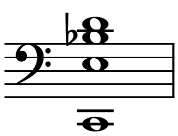
where the function `round` maps a real number to the nearest integer. The max function is used instead of a sum in order to account for the masking effect (Moore, 1982). The quantization given by the rounding function corresponds to the fact that most of the tonal music is composed using the *well-tempered tuning*. For instance, the 3rd harmonic $f_{3,i}$ corresponds to a note $i+7$ which is located one perfect fifth (i.e. 7 semi-tones) over the note i corresponding to the fundamental frequency. Building the whole set of possible notes from that principle leads to a system where flat and sharp notes are not the same, which was found to be impractical by musical instrument designers in the baroque era. Since then, most Western musicians used a compromise called the well-tempered scale, where semi-tones are separated by an equal ratio of frequencies. Hence, the rounding function in Equation (1) provides a frequency quantization that corresponds to what an average contemporary music listener experiences on a regular basis.

For each chord $\mathcal{X}_j$, we then have a distributed representation $\mathbf{l}_j = \{l_j(n_1), \ldots, l_j(n_d)\}$ corresponding to the perceived strength of the harmonics related to every note n_k of the well-tempered scale, where we consider the d first notes of this scale to be relevant. For instance, one can set the range of the notes n_1 to n_d to correspond to audible frequencies. Using octave invariance, we can go further and define a chord representation $\mathbf{v}_j = \{v_j(0), \ldots, v_j(11)\}$ where

$$v_j(i) = \sum_{n_k : 1 \leq k \leq d,\ (n_k \bmod 12) = i} l(n_k). \tag{2}$$

This representation gives a measure of the relative strength of each pitch class[3] in a given chord. For instance, value $v_j(0)$ is associated with pitch class `c`, value $v_j(1)$ to pitch class `c` sharp, and so on.

Throughout this paper, we define chords by giving the pitch class letter, sometimes followed by symbol # (sharp) to raise a given pitch class by one semi-tone. Finally, each pitch class is followed by a digit representing the actual octave where the note is played. For instance, the symbol `c1e2a#2d3` stands for the 4-note chord



with a `c` on the first octave, an `e` and an `a` sharp (`b` flat) on the second octave, and finally a `d` on the third octave.

[3]All notes with the same note name (e.g. C#) are said to be part of the same *pitch class*.

Table 1: Euclidean distances between the chord in the first row and other chords when chord representation is given by Equation (2), choosing $\rho = 0.97$.

c1a2e3g3	0.000	c1d#2a#2d3	0.000
c1a2c3e3	1.230	c1a#2d#3g3	1.814
c1a2d3g3	1.436	c1e2a#2d#3	2.725
c1a1d2g2	2.259	c1a#2e3g#3	3.442
c1a#2e3a3	2.491	c1e2a#2d3	3.691
a0c3g3b3	2.920	a#0d#2g#2c3	3.923
c1e2b2d3	3.162	a#0d2g#2c3	4.155
c1g2c3e3	3.398	g#1g2c3d#3	4.363
a0g#2c3e3	3.643	c1e2a#2c#3	4.612
c1f2c3e3	3.914	a#1g#2d3g3	4.820
c1d#2a#2d3	4.295	f1a2d#3g3	5.030
e1e2g2c3	4.548	d1f#2c3f3	5.267
g1a#2f3a3	4.758	a0c3g3b3	5.473
e0g2d3f#3	4.969	g1f2a#2c#3	5.698
f#0e2a2c3	5.181	b0d2a2c3	5.902
g#0g2c3d#3	5.393	e1d3g3b3	6.103
f#1d#2a2c3	5.601	f#1e2a#2d#3	6.329
g0f2b2d#3	5.818	d#1c#2f#2a#2	6.530
g1f2a#2c#3	6.035	g#0b2f3g#3	6.746
g1f2b2d#3	6.242	b0a2d#3g3	6.947

We see in Figure 1 that the representation given by Equation (2) provides similar results for two different voicings of the C major chord, as defined in Levine (1990). We have also computed Euclidean distances between chords induced by this representation and found that they roughly correspond to perceptual closeness, as shown in Table 1. Each column gives Euclidean distances between the chord in the first row and some other chords that are represented as described here. The trained musician should see that these distances roughly correspond to perceived closeness. For instance, the second column is related to a particular inversion of the C minor chord (`c1d#2a#2d3`). We see that the closest chord in the dataset (`c1a#2d#3g3`) is the second inversion of the same chord, as described in Levine (1990). Hence, we raise the note d#3 by one octave and replace the note d3 by g3 (separated by a perfect fourth). These two notes share some harmonics, leading to a close vectorial representation. This distance measure could have considerable interest in a broad range of computational generative models in music as well as for music composition.

2.2 Probabilities of Substitution

As already pointed out in Section 2, there is no direct way to represent Euclidean distances between discrete objects in the graphical model framework. Considering the particular problem of chord progressions, it is however possible to convert the Euclidean distances described in section 2.1 into probabilities of substitution between chords in a given corpus of chord progression.

One can define the probability $p_{i,j}$ of substituting

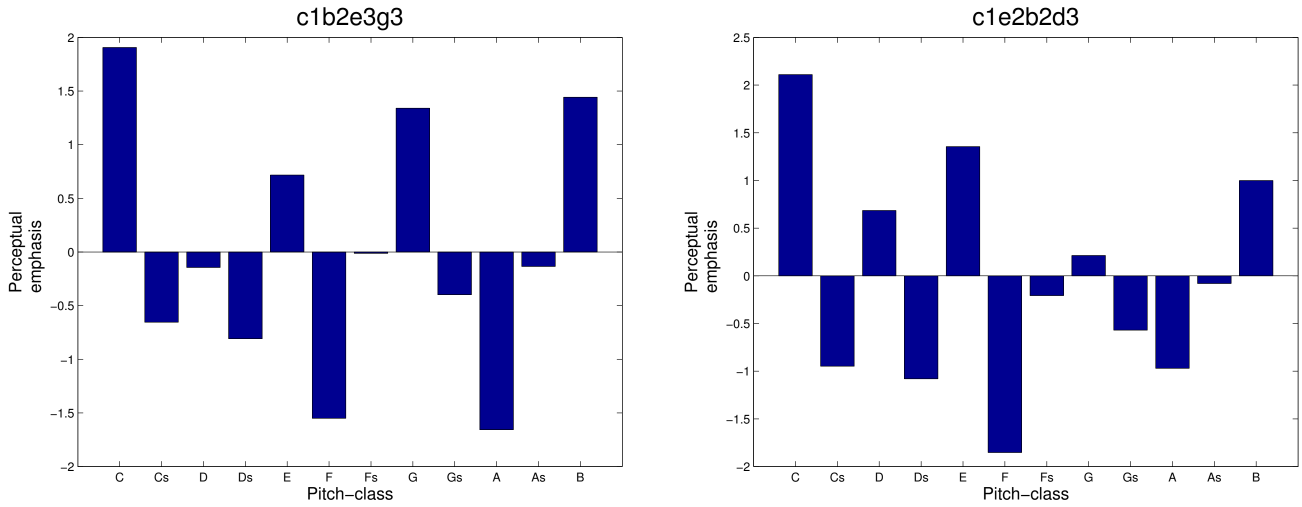


Figure 1: Normalized values given by Equation (2) for 2 voicings of the C major chord. We see that perceptual emphasis is higher for pitch classes present in the chord. These two chord representations have similar values for pitch classes that are not present in either chords, which makes their Euclidean distance small.

chord $\mathcal{X}_i$ for chord $\mathcal{X}_j$ in a chord progression as

$$p_{i,j} = \frac{\phi_{i,j}}{\sum_{1 \leq j \leq s} \phi_{i,j}} \quad (3)$$

with

$$\phi_{i,j} = \exp\{-\lambda ||\mathbf{v}_i - \mathbf{v}_j||^2\}$$

with free parameter $0 \leq \lambda < \infty$. The parameters λ and ρ (from Equation (1)) can be optimized by validation on any chord progression dataset provided a suitable objective function. With possible values going from 0 to arbitrary high values, the parameter λ allows the substitution probability table to go from the uniform distribution with equal entries everywhere (such that every chord has the same probability of being played) to the identity matrix (which disallow any chord substitution). Table 2 shows substitution probabilities obtained from Equation (3) for chords in Table 1.

3 Graphical Model

Graphical models (Lauritzen, 1996) are a useful framework to describe probability distributions where graphs are used as representations for a particular factorization of joint probabilities. Vertices are associated with random variables. If two vertices are not linked by an edge, their associated random variables are considered to be unconditionally independent. A directed edge going from the vertex associated with variable A to the one corresponding to variable B accounts for the presence of the term $P(B|A)$ in the factorization of the joint distribution for all the variables in the model. The process of calculating probability distributions for a subset of the variables of the model given the joint distribution of all the variables is called *marginalization* (e.g. deriving $P(A, B)$ from $P(A, B, C)$). The graphical model framework provides efficient algorithms for marginalization and various learning algorithms can be used to learn the parameters of a model, given an appropriate dataset.

We now propose a graphical model for chord sequences using the probabilities of substitution between chords described in Section 2.2. The main assumption behind the proposed model is that conditional dependencies between chords in a typical chord progression are strongly tied to the metrical structure associated with it. Another important aspect of this model is that it is not restricted to local dependencies, like a simpler Hidden Markov Model (HMM) (Rabiner, 1989) would be. This choice of structure reflects the fact that a chord progression is seen in this model as a two dimensional architecture. Every chord in a chord progression depends both on its position in the chord structure (global dependencies) and on the surrounding chords (local dependencies). We show empirically in Section 4 that considering both aspects leads to better generalization performance as well as better generated results than by only considering local dependencies.

Variables in a graphical model can be either observed in the dataset or hidden (i.e. not present in the dataset but still present in the model). The Expectation-Maximization (EM) algorithm (Dempster et al., 1977) can be used to estimate the conditional probabilities of the hidden variables of a probabilistic model. This algorithm proceed in two steps applied iteratively to each observations in a dataset until convergence of the parameters. First, the E step compute the expectation of the hidden variables, given the current parameters of the model and one observation in the dataset. Secondly, the M step update the values of the parameters in order to maximize the likelihood of the same observation combined to the expected values for the hidden variables.

Figure 2 shows a graphical model that can be used as a generative model for chord progressions in this fashion. All the random variables in the model are discrete. Nodes in level 1, 2 and 3 are hidden while nodes in level 4 are observed. Every chords are represented as distinct discrete

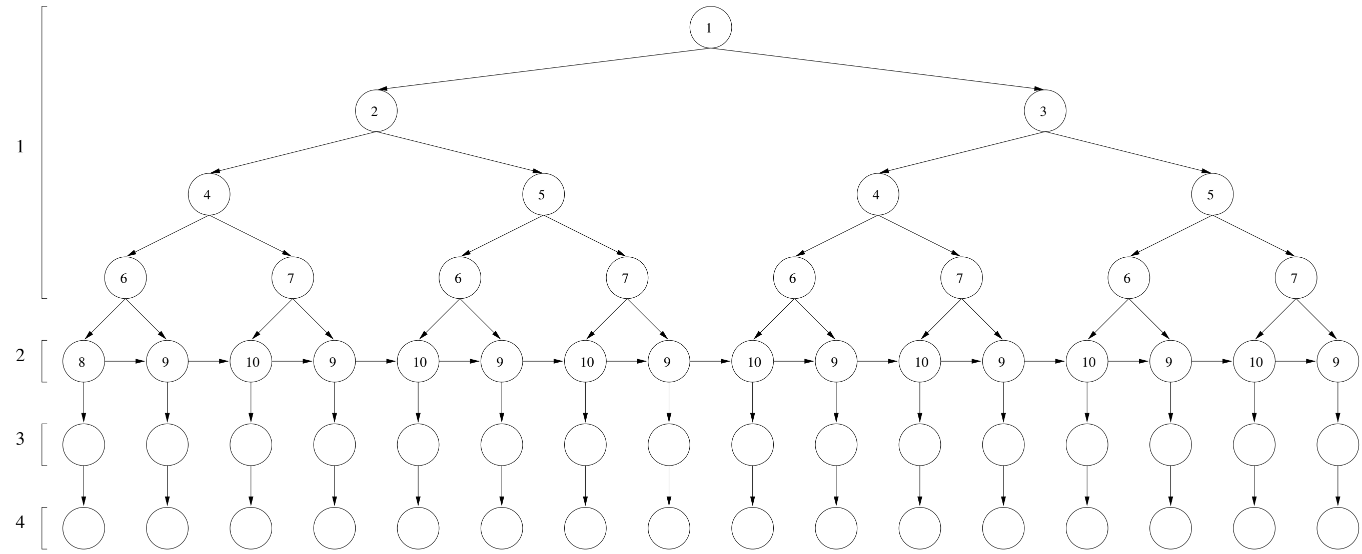


Figure 2: A probabilistic graphical model for chord progressions, as described in Section 3. Numbers in level 1 and 2 nodes indicate a particular form of parameter sharing that has been used in the experiments (see Section 4.1).

events. Nodes in level 1 directly model the contextual dependencies related to the meter. Nodes in level 2 combine this information with local dependencies in order to model smooth chord progressions. Variables in level 1 and 2 have an arbitrary number of possible states optimized by cross-validation (Hastie et al., 2001). The upper tree structure makes it possible for the algorithm to generate proper endings. Smooth voice leadings (e.g. small distances between generated notes in two successive chords) are made possible by the horizontal links in level 2.

Variables in levels 3 and 4 have a number of possible states equal to the number of chords in the dataset. Hence, each state is associated with a particular chord. The probability table associated with the conditional dependencies going from level 3 to 4 is fixed during learning with the values given by Equation (3). Values in level 3 are hidden and represent intuitively "initial" chords that could have been substituted by the actual observed chords in level 4.

The role of the fixed substitution matrix is to raise the probability of unseen events in a way that account for psychoacoustical similarities. Discarding level 4 and directly observing nodes in level 3 would assign extremely low probabilities to unseen chords in the training set. Instead, when observing a given chord on level 4 during learning, the probabilities of *every* chords of the dataset are updated with respect to the probabilities of substitution described in Section 2.2.

The marginalization in the graphical model can be achieved using the Junction Tree Algorithm (JTA) (Lauritzen, 1996). By defining a convenient factorization of all the variables from the one defined by the graph, the JTA allows marginalization of small subsets of the variables to be done efficiently. Exact marginalization techniques are tractable in this model given its limited complexity.

Many variations of this particular model are possible, some of which are compared in Section 4. We say that two random variables are "tied" when they share the same conditional probability parameters. Conditional probability tables in the proposed model can be tied in various ways. Also, more horizontal links in the model can be added to reinforce the dependencies between higher level hidden variables.

Other tree structures may be more suitable for music having different meters (e.g. ternary structures for waltzes). Using a tree structure has the advantage of reducing the complexity of the considered dependencies from the order m to the order $\log m$, where m is the length of a given chord sequence. It should be pointed out that in this paper we only consider musical productions with fixed length. Fortunately, the current model could be easily extended to variable length musical production by adding conditional dependencies arrows between many normalized subtrees.

4 Experiments

52 jazz standards excerpts from Sher (1988) were interpreted and recorded by the first author in MIDI format on a Yamaha Disklavier piano. See `http://www.idiap.ch/~paiement/chords/` for a listing. Standard 4-note jazz piano voicings as described in Levine (1990) were used to convert the chord symbols into musical notes. Thus, the model is considering chord progressions as they might be expressed by a trained jazz musician in a realistic musical context. The complexity of the chord sequences found in the corpus is representative of the complexity of common chord progressions in most jazz and pop music. We chose to record actual voiced chords rather than symbolic chord names (e.g. *Em7*) because the symbolic names are ineffective at capturing the specific voicings made by a trained jazz musician.

Every jazz standard excerpt was 16 bars long, with a 4 beat meter, and with one chord change every 2 beats (yielding observed sequences of length 32). Longer chords were repeated multiple times (e.g. a 6 beat chord

Table 2: Subset of the substitution probability table constructed with Equation (3). For each column, the number in the first row corresponds to the probability of playing the associated chord with no substitution. The numbers in the following rows correspond to the probability of playing the associated chord instead of the chord in the first row of the same column.

c1a2e3g3	0.41395	c1d#2a#2d3	0.70621
c1a2c3e3	0.08366	c1a#2d#3g3	0.06677
c1a2d3g3	0.06401	c1e2a#2d#3	0.02044
c1a1d2g2	0.02195	c1a#2e3g#3	0.00805
c1a#2e3a3	0.01623	c1e2a#2d3	0.00582
a0c3g3b3	0.00929	a#0d#2g#2c3	0.00431
c1e2b2d3	0.00679	a#0d2g#2c3	0.00318
c1g2c3e3	0.00500	g#1g2c3d#3	0.00243
a0g#2c3e3	0.00363	c1e2a#2c#3	0.00176
c1f2c3e3	0.00255	a#1g#2d3g3	0.00134
c1d#2a#2d3	0.00156	f1a2d#3g3	0.00102
e1e2g2c3	0.00112	d1f#2c3f3	0.00075
g1a#2f3a3	0.00085	a0c3g3b3	0.00057
e0g2d3f#3	0.00065	g1f2a#2c#3	0.00043
f#0e2a2c3	0.00049	b0d2a2c3	0.00033
g#0g2c3d#3	0.00037	e1d3g3b3	0.00025
f#1d#2a2c3	0.00028	f#1e2a#2d#3	0.00019
g0f2b2d#3	0.00021	d#1c#2f#2a#2	0.00015
g1f2a#2c#3	0.00016	g#0b2f3g#3	0.00011
g1f2b2d#3	0.00012	b0a2d#3g3	0.00008

is represented as 3 distinct 2-beat observations). This simplification has a limited impact on the quality of the model since generating a chord progression is simply a first (but very important) step toward generating complete polyphonic music, where modeling actual event lengths would be more crucial. The jazz standards were carefully chosen to exhibit a 16 bars global structure. We used the last 16 bars of each standard to train the model. Since every standard ends with a cadenza (i.e. a musical ending), the chosen excerpts exhibit strong regularities.

4.1 Generalization

The chosen discrete chord sequences were converted into sequences of 12-dimensional continuous vectors as described in Section 2. Frequencies ranging from 30Hz to 20kHz (MIDI notes going from the lowest note in the corpus to note number 135) were considered in order to build the representation given by Equation (1). It is possible to measure how well a given architecture captures conditional dependencies between sub-sequences. In order to do so, average negative *conditional* out-of-sample likelihoods of sub-sequences of length 8 on positions 1, 9, 17 and 25 have been computed. For each sequence of chords $\mathbf{x} = \{x_1, \ldots x_{32}\}$ in the appropriate validation set, we average the values

$$-\log P(x_i, \ldots, x_{i+7} | x_1, \ldots, x_{i-1}, x_{i+8}, \ldots, x_{32}).$$

with $i \in \{1, 9, 17, 25\}$. Hence, the likelihood of each subsequence is conditional on the rest of the sequence (taken in the validation set) from which it originates. Double cross-validation is a recursive application of cross-validation where both the optimization of the parameters of the model and the evaluation of the generalization of the model are carried out simultaneously. This technique has been used to optimize the number of possible values of hidden variables and the parameters ρ and λ for various architectures. Results are given in Table 3.

Table 3: Average negative conditional out-of-sample log-likelihoods of sub-sequences of length 8 on positions 1, 9, 17 and 25, given the rest of the sequences. These results are computed using double cross-validation in order to optimize the number of possible values for hidden variables and the parameters λ and ρ. We see that the trees perform better than the HMM.

Model	(Tying in level 1)	Negative log-likelihood
Tree	No	32.3281
Tree	Yes	32.6364
HMM		33.2527

Two forms of parameter tying for the tree model have been tested. The conditional probability tables in level 1 of Figure 2 can be either tied as shown by the numbers inside the nodes in the figure or can be left untied. Tying for level 2 is always done as illustrated in Figure 2 by the numbers inside the nodes, to model local dependencies. All nodes in level 3 share the same parameters for all tested models. Also, recall that parameters for the conditional probabilities of variables in level 4 are fixed as described in Section 3.

As a benchmark, an HMM consisting of levels 2, 3 and 4 of Figure 2 has been trained and evaluated on the same dataset. The results presented in Table 3 are similar to perplexity or prediction ability. We choose this particular measure of generalization to account for the binary metrical structure of the chord progressions in the corpus. The fact that these contextual out-of-sample likelihoods are better for the trees than for the HMM are an indication that time-dependent regularities are present in the data. Further investigations would be necessary in order to assess to what extent chord structures are hierarchically related to the meter.

4.2 Generation

One can sample the joint distribution learned by the proposed model in order to generate novel chord progressions. Chord progressions generated by the models presented in this paper are available at `http://www.idiap.ch/~paiement/chords/`. For instance, Figure 3 shows a chord progression that has been generated by the graphical model shown in Figure 2. This chord progression has all the characteristics of a standard jazz chord progression. A trained musician may observe that the last 8 bars of the sequence is a II-V-I[4] chord progression (Levine, 1990), which is very common.

[4] The lowest notes are d, g and c.

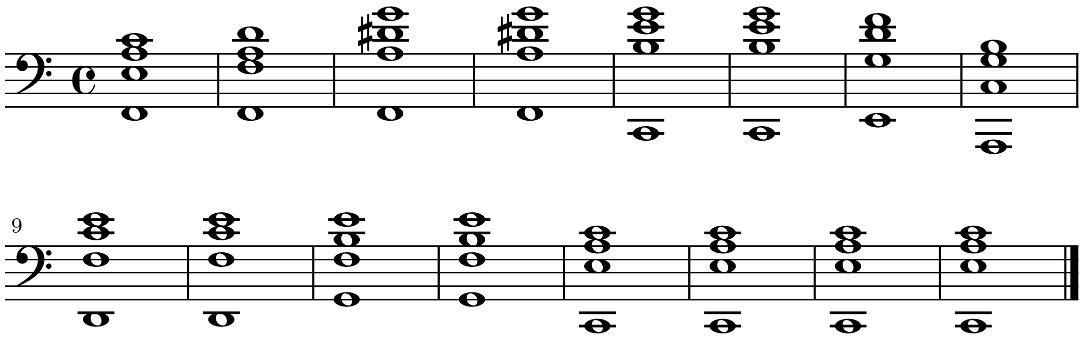


Figure 3: A chord progression generated by the proposed model. This chord progression is very similar to a standard jazz chord progression.

For comparison Figure 4 shows a chord progression generated by the HMM model. While the chords follow one another in a smooth fashion, there is no global coherence among the chords. For instance, one can see that the lowest note of the last chord is not a C, which was the case for all the chord sequences in the training set. The fundamental qualitative difference between both methods should be obvious even for the non-musician when listening to the generated chord sequences.

5 Conclusion

We have shown empirically that chord progressions exhibit global dependencies that can be better captured with a tree structure related to the meter than with a simple dynamical HMM that concentrates on local dependencies. The importance of contextual information for modeling chord progressions is even more apparent when one compares sequences of chords sampled from both models. The time-dependent hidden variables enable the tree structure to generate coherent chord progressions both locally and globally.

However, the low difference in terms of conditional out-of-sample likelihood between the tree model and the HMM, and the relatively low number of degrees of freedom for optimal generalization (including the low optimal number of possible state for hidden variables) are a good indication that increasing the number of sequences in the dataset would probably be necessary in further developments of probabilistic models for chord progressions. Also, a better evaluation of such models could be achieved by using them for a supervised task. Applications where a chord progression model could be included range from music transcription, music information retrieval, musical genre recognition to music analysis applications, just to name a few.

Chord progressions are regular and simple structures that condition dramatically the actual choice of notes in polyphonic tonal music. Hence, we argue that chord models are crucial in the design of efficient algorithms that deal with such music data. Moreover, generating interesting chord progressions may be one of the most important aspects in generating realistic polyphonic music. Our model constitutes a first step in that direction.

ACKNOWLEDGEMENTS

The first author would like to thank Yves Grandvalet and David Barber for helpful discussions. This work was supported in part by the IST Program of the European Community, under the PASCAL Network of Excellence, IST-2002-506778, funded in part by the Swiss Federal Office for Education and Science (OFES) and the Swiss NSF through the NCCR on IM2.

References

M. Allan and C. K. I. Williams. Harmonising chorales by probabilistic inference. In *Advances in Neural Information Processing Systems*, volume 17, 2004.

Y. Bengio, P. Simard, and P. Frasconi. Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 5(2):157–166, 1994.

A. T. Cemgil. *Bayesian Music Transcription*. PhD thesis, Radboud University of Nijmegen, 2004.

Grosvenor Cooper and Leonard B. Meyer. *The Rhythmic Structure of Music*. The Univ. of Chicago Press, 1960.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society*, 39:1–38, 1977.

Douglas Eck and Juergen Schmidhuber. Finding temporal structure in music: Blues improvisation with LSTM recurrent networks. In H. Bourlard, editor, *Neural Networks for Signal Processing XII, Proc. 2002 IEEE Workshop*, pages 747–756, New York, 2002. IEEE.

Stephen Handel. *Listening: An introduction to the perception of auditory events*. MIT Press, Cambridge, Mass., 1993.

T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer series in statistics. Springer-Verlag, 2001.

http://cf.geocities.com/chordprogs. Temporary anonymous URL to put audio samples generated by the proposed algorithms. A proper academic website will be established if the paper is accepted., 2005.

T. Kuusi. *Set-Class and Chord: Examining Connection Between Theoretical Ressemblance and Perceived*

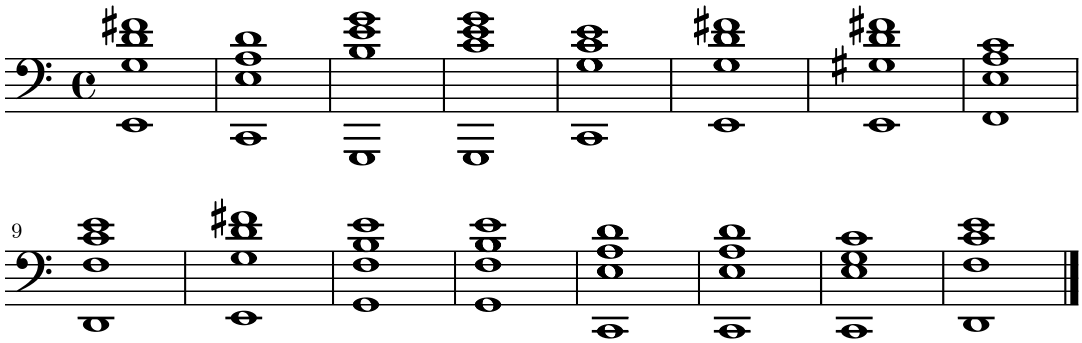


Figure 4: A chord progression generated by the HMM model. While the individual chord transitions are smooth and likely, there is no global chord structure.

Closeness. Number 12 in Studia Musica. Sibelius Academy, 2001.

S. L. Lauritzen. *Graphical Models*. Oxford University Press, 1996.

V. Lavrenko and J. Pickens. Polyphonic music modeling with random fields. In *Proceedings of ACM Multimedia*, Berkeley, CA, November 2-8 2003.

Mark Levine. *The Jazz Piano Book*. Sher Music Co./Advance Music, 1990.

B.C.J. Moore. *An Introduction to the Psychology of Hearing*. Academic Press, 1982.

L. R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–285, February 1989.

C. Raphael and J. Stoddard. Harmonic analysis with probabilistic graphical models. In *Proceedings of ISMIR 2003*, 2003.

Chuck Sher, editor. *The New Real Book*, volume 1. Sher Music Co., 1988.

V. Valimaki, J. Huopaniemi, Karjaleinen, and Z. Janosy. Physical modeling of plucked string instruments with application to real-time sound synthesis. *J. Audio Eng. Society*, 44(5):331–353, 1996.

P. Vassilakis. Chords as spectra, harmony as timbre. *J. Acoust. Soc. Am.*, 106(4/2):2286, 1999. (paper presented at the 138th meeting of the Acoustical Society of America).