A Graphical Model for Chord Progressions Embedded in a Psychoacoustic Space

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

Douglas Eck ECKDOUG@IRO.UMONTREAL.CA
Dep. of Computer Science and Operations Research, University of Montreal, CP 6128 succ. Centre-Ville, Montreal, Canada, H3C 3J7

Samy Bengio BENGIO@IDIAP.CH
David Barber BARBER@IDIAP.CH
IDIAP Research Institute, Rue du Simplon 4, Case Postale 592, CH-1920 Martigny, Switzerland

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. Parameters in the graphical models are learnt with the EM algorithm and the classical Junction Tree algorithm. 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.

Tonal music comprises most of the music that has been written since J.-S. Bach (including contemporary pop music.) 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.

The interaction between the notes that are actually played and the notes comprising the chord progression are related to the meter of the piece. Meter is the sense of strong and weak beats that arises from the interaction among hierarchical levels of sequences having nested periodic components. 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, the first and third beats are usually emphasized. In terms of melodic structure, this indicates that notes perceptually closer to the chord progression are more likely to

Appearing in *Proceedings of the 22^{nd} International Conference on Machine Learning*, Bonn, Germany, 2005.

be played on these beats while more "dissonant" notes can be played on weaker beats.

The strong link between chord structure and overall structure motivate 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 automatically generates melodies. As an example, consider blues music. Most blues compositions are variations of a *single* 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 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 composer (Eck & 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 (2005) 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.

The first contribution in this paper is the introduction in Section 2 of a similarity measure for chords guided by psychoacoustic considerations. A probabilistic model for chord progressions taking into account long term dependencies is then described in Section 3 as the other main contribution in this paper. 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. Representation

The generalization performance of a generative model depends strongly on how observed data is represented. A good representation encapsulates some of the psychoacoustic similarities between chords. 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 are seen as a single timbre (Vassilakis, 1999) and we

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

[2] MIDI stands for Musical Instrument Digital Interface, an industry-standard interface used on electronic musical keyboards and PCs for computer control of musical instruments and devices. In our work, we only consider notes onsets and offsets in the MIDI signal.

design a continuous distributed representation where close chords with respect to Euclidean distance tend to be similar to listeners.

More specifically, 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 $m(f)$. Then, for a given chord, we associate to each MIDI note n a perceived loudness $l(n)$ with

$$l(n) = \max(\{\rho^h | \texttt{round}(m(f_{h,i})) = n\} \cup \{0\}) \qquad (1)$$

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 which is located one perfect fifth (i.e. 7 semi-tones) over the note 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, musicians used a compromise called the well-tempered scale, where semitones 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, we then have a distributed representation $\mathbf{l} = \{l(n_1), \ldots, l(n_d)\}$ corresponding to the perceived strength of the harmonics related to every note n_i of the well-tempered scale. Using octave invariance, we can go further and define a chord representation $\mathbf{v} = \{v(0), \ldots, v(11)\}$ where

$$v(i) = \sum_{n_j : 1 \leq j \leq d,\ (n_j \bmod 12) = i} l(n_j). \qquad (2)$$

This representation gives a measure of the relative strength of each pitch class[3] in a given chord. For instance, value $v(0)$ is associated to pitch class `c`, value $v(1)$ to pitch class `c#`, and so on[4]. We see in Figure 1 that this representation gives similar results for two different voicings of the C major chord, as defined in Levine (1990). By implicitly quantifying psychoacoustic similarities between chords, such a low dimensional representation has considerable interest for developping complex probabilistic models for chords progressions and is thus a main contribution of our work.

[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).

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
a0g#2c3e3	3.643	c1e2a#2c#3	4.612
c1d#2a#2d3	4.295	f1a2d#3g3	5.030
g1a#2f3a3	4.758	a0c3g3b3	5.473
f#0e2a2c3	5.181	b0d2a2c3	5.902
f#1d#2a2c3	5.601	f#1e2a#2d#3	6.329
g1f2a#2c#3	6.035	g#0b2f3g#3	6.746

We have computed Euclidean distances between chords represented in the distributed chord space given by Equation (2) 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#2` by one octave and replace the note `d3` by `g3` (separated by a perfect fourth.) These two notes are sharing 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.

[4]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 a 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.

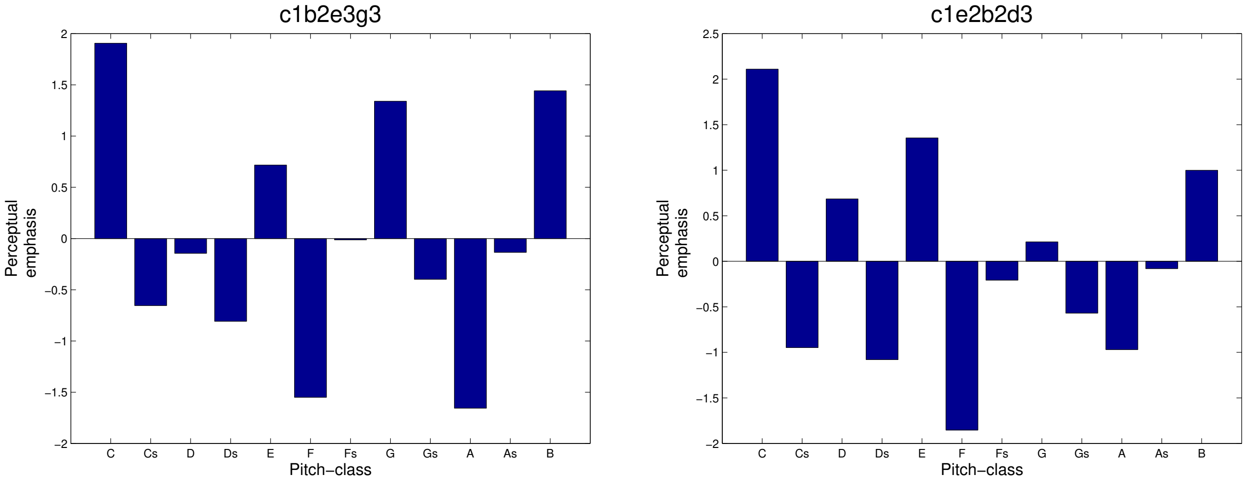

Figure 1. Normalized values given by Equation (2) for 2 voicings of the C major chord. We see that perceptual emphasis is higher for pitch classes present in the chord. These two chord representations have similar values for pitch classes that are not present in either chords, which makes their Euclidean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

3. Graphical Model

We now propose a graphical model that generates chord sequences using the input representation described in Section 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 to it. Another important aspect of this model is that it is not restricted to local dependencies, like a simpler Hidden Markov Model (HMM) 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 in Section 4 that considering both aspects leads to better generalization performance as well as better generated results than by only considering local dependencies.

The design of our model is motivated by theories of musical rhythm (Cooper & Meyer, 1960) and music structure (Lerdahl & Jackendoff, 1983). A given musical note does not itself have a certain meaning. Its meaning, if any, is defined by the role it plays in longer musical elaborations such as melodies. To make an analogy to language, musical notes are perhaps more similar to *letters* than to *words*. However, the analogy is not entirely correct because even musical phrases do not have meaning in isolation in the same way that words do. A principal source of music structure is the meter of a piece. Almost all Western music is metered, indicating a fixed hierarchical temporal structure with small integer relationships between levels. We used meter to guide the construction of probabilistic trees, employing a binary tree structure suggested by the meter of the jazz standards in our database. Though this tree structure differs from that of other forms of music (thus representing a built-in stylistic prior motivated by music theory) the difference is not as great as it might seem. Most meters yield binary trees similar to the one we employ. Furthermore, if a tree is non-binary, then it is usually so only on a single level. For example, in a typical 3/4 piece of waltz music, the quarter-note level is indeed ternary (3:1). However, the higher-level relationships remain binary, with musical phrases being formed out of 2, 4 or 8 measures.

Figure 2 shows a graphical model constructed as described above. Discrete nodes in levels 1 and 2 are not observed. The purpose of the nodes in level 1 is to capture global chord dependencies related to the meter. Nodes in level 2 are modeling local chord dependencies conditionally to the global dependencies captured in level 1. For instance, the fact that the algorithm is accurately generating proper endings is constrained by the upper tree structure. On the other hand, the smoothness of the voice leadings (e.g. small distances between generated notes in two successive chords) is modeled by the horizontal links in level 2.

The bottom nodes of the model are continuous observations conditioned by discrete hidden variables. Hence, a mixture of Gaussians can be used to model each observation given by the distributed representation described in Section 2. Suppose a Gaussian node G has a discrete parent D, then the conditional density $p(G|D)$ is given by

$$p(G|D = i) \sim \mathcal{N}(\mu_i, \sigma_i) \quad (3)$$

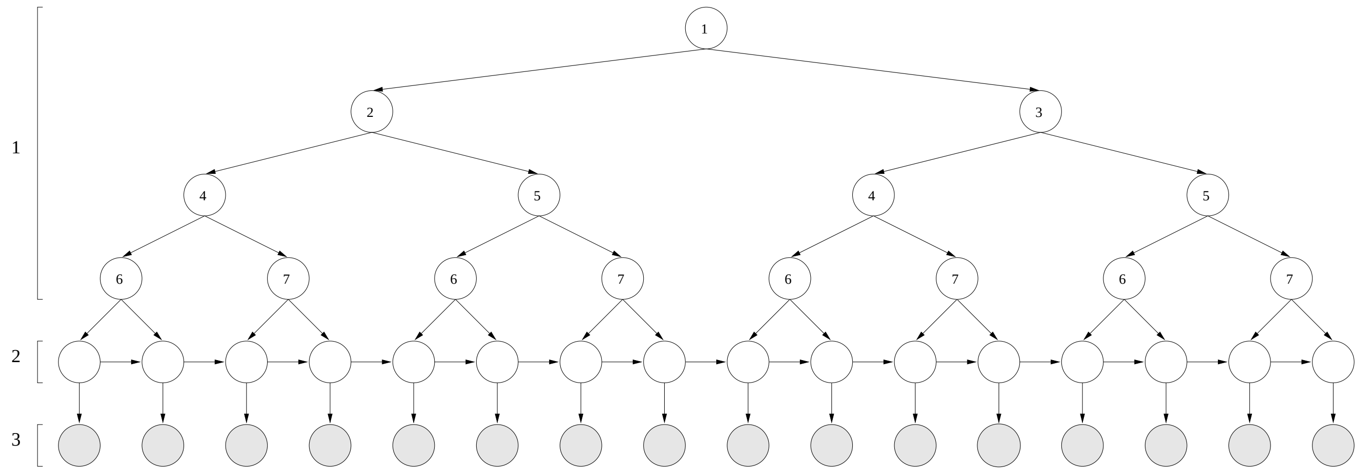

Figure 2. A probabilistic graphical model for chord progressions. White nodes correspond to discrete hidden variables while gray nodes correspond to observed multivariate mixtures of Gaussians nodes. Nodes in level 1 directly model the contextual dependencies related to the meter. Nodes in level 2 combine this information with local dependencies in order to model smooth chord progressions. Finally, continuous nodes in level 3 are observing chords embedded in the continuous space defined by Equation (2). Numbers in level 1 nodes indicate a particular form of parameter sharing that has proven to be useful for generalization (see Section 4.1).

where $\mathcal{N}(\mu, \sigma)$ is a k-dimensional Gaussian distribution with mean $\mu \in \mathbb{R}^k$ and diagonal covariance matrix $\Sigma \in \mathbb{R}^k \times \mathbb{R}^k$ determined by its diagonal elements $\sigma \in \mathbb{R}^k$.

The Expectation-Maximization (EM) algorithm (Dempster et al., 1977) can be used to estimate the conditional probabilities of the hidden variables in a graphical model. This algorithm proceeds in two steps applied iteratively over a dataset until convergence of the parameters. First, the E step computes the expectation of the hidden variables, given the current parameters of the model and the observations of the dataset. Secondly, the M step updates the values of the parameters in order to maximize the likelihood of the observations using the expected values of the hidden variables.

Marginalization must be carried out in the proposed model both for learning (during the expectation step of the EM algorithm) and for evaluation. The inference in a graphical model can be achieved using the Junction Tree Algorithm (JTA) (Lauritzen, 1996). In order to build the junction tree representation of the joint distribution of all the variables of the model, we start by moralizing the original graph (i.e. connecting the non-connected parents of a common child and then removing the directionality of all edges) so the independence properties in the original graph are preserved. In the next step (called triangulation), we add edges to remove all chord-less cycles of length greater than 4. Finally, we can form clusters with the maximal cliques of the triangulated graph. The junction tree representation is formed by joining these clusters together. To each cluster, we associate a potential function which can be normalized to give the marginalized probabilities of the variables in that cluster. Given evidence, the properties of the junction tree allow these potential functions to be updated by local message passing. Exact marginalization techniques are tractable in the proposed model given its limited complexity.

Many variations of the proposed graphical structure are possible, some of which are compared in Section 4. For instance, conditional probability tables 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. The chord progressions are intimately tied to the metrical structure, which has obviously binary structure in the corpus of data. However, 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 chords sequences with variable length by adding conditional dependencies arrows between many normalized subtrees.

Considering global dependencies to model time series is a general issue also present in other domains. For instance, tree models with structures derived from common syntactical patterns could be used to learn global dependencies in natural language processing applications. However, it should be noted that dependencies are much more complex in natural language than in chord progressions.

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/icml2005` 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 8 bars long, with a 4 beats meter, and with one chord change every 2 beats (yielding observed sequences of length 16.) Longer chords were repeated multiple times (e.g. a 6 beats chord 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 8 bars of each standards to train the model. Since every standard ends with a cadenza (i.e. a musical ending), the chosen excerpts exhibits 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 20Hz 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). A value of ρ of 0.96 was arbitrarily chosen for the experiments. It should be pointed out that since the generative models have been trained in an unsupervised setting, it is irrelevant to compare different chord representations in terms of likelihood. However, it is possible to measure how well a given architecture is modeling conditional dependencies between sub-sequences of chords. In order to do so, average negative conditional out-of-sample likelihoods of sub-sequences of length 4 on positions 1, 5, 9 and 13 have been computed. 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 has been used to optimize the number of possible values of hidden variables for various architectures. Results are given in Table 2. This measure is similar to perplexity or prediction ability. We chose this particular measure of generalization in order to account for the binary metrical structure of chord progressions, which is not present in natural language processing, for instance.

Table 2. Average conditional negative out-of-sample log-likelihoods of sub-sequences of length 4 on positions 1, 5, 9 and 13. These results are computed using double cross-validation in order to optimize the number of possible values for hidden variables. The numbers in parentheses indicate which levels of the tree are tied as described in Section 4.1. We see that some combinations of parameter tying in the trees performs better than the standard HMM.

Model (tying)	Negative log-likelihood
Tree (2, 3)	93.8910
Tree (1, 3)	94.0037
Tree (1, 2, 3)	94.9309
Tree (3)	98.2446
HMM	98.2611

Different forms of parameter tying for the tree model shown in Figure 2 have been tested. All nodes in level 3 share the same parameters for all tested models. Hence, we use only one 12-dimensional mixture of Gaussians (as in Equation (3)) independently of time, in order to constrain the capacity of the model. Moreover, a diagonal covariance matrix Σ has been used, thus reducing the number of free parameters to 24 in level 3 (12 for μ and 12 for Σ). Hidden variables in level 1 and 2 can be tied or not. Tying for level 1 is done as illustrated in Figure 2 by the numbers inside the nodes.

The fact that the contextual out-of-sample likelihoods presented in Table 2 are better for the different trees than for the HMM indicates that time-dependent regularities are present in the data. Sharing parameters in levels 1 or 2 of the tree increases the out-of-sample likelihood. This indicates that regularities are repeated over time in the signal. Further investigations would be necessary in order to assess to what extent chord structures are hierarchically related to the meter. On the other hand, the relatively high values obtained in terms of conditional out-of-sample negative log-likelihood indicates that the number of sequences may not be sufficient to efficiently represent the variability of the data. Unfortunately, reliable chord progressions data is difficult to generate. However, the authors plan to extend the chord database significantly in the short term.

4.2. Generation

One can sample the proposed model in order to generate novel chord progressions. Fortunately, Euclidean distances are relevant in the observation space created in Section 2. Thus, a simple approach to generate chord progressions is to take the nearest neighbors (nearest chords in the training set) of each sampled values obtained by sampling the observation nodes.

Chord progressions generated by the models presented in this paper are available at `http://www.idiap.ch/~paiement/icml2005`. 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. For instance, the trained musician can observe that the last 8 bars of the sequence is a II-V-I[5] chord progression (Levine, 1990), which is very common. Figure 4 shows a chord progression generated by the HMM model. While the chords are following each other in a smooth fashion, there is no global relation between 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

In this paper, we introduced a low dimensional representation for chords that can be used as observations for probabilistic models. This constitutes the first main contribution in our work. Moreover, 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 states 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 including 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.

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.

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

Acknowledgements

The first author would like to thank Yves Grandvalet 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

Allan, M., & Williams, C. K. I. (2005). Harmonising chorales by probabilistic inference. *Advances in Neural Information Processing Systems*.

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

Cemgil, A. T. (2004). *Bayesian music transcription*. Doctoral dissertation, Radboud University of Nijmegen.

Cooper, G., & Meyer, L. B. (1960). *The Rhythmic Structure of Music*. The Univ. of Chicago Press.

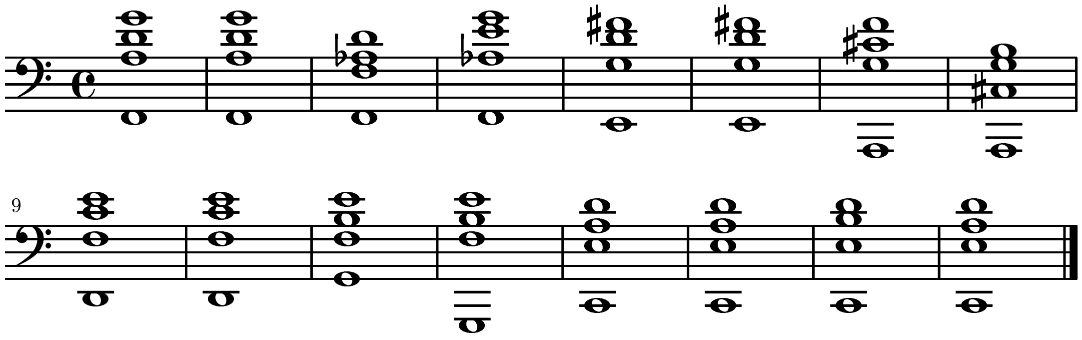

Figure 3. A chord progression generated by the proposed model. This chord progression is very similar to a standard jazz chord progression.

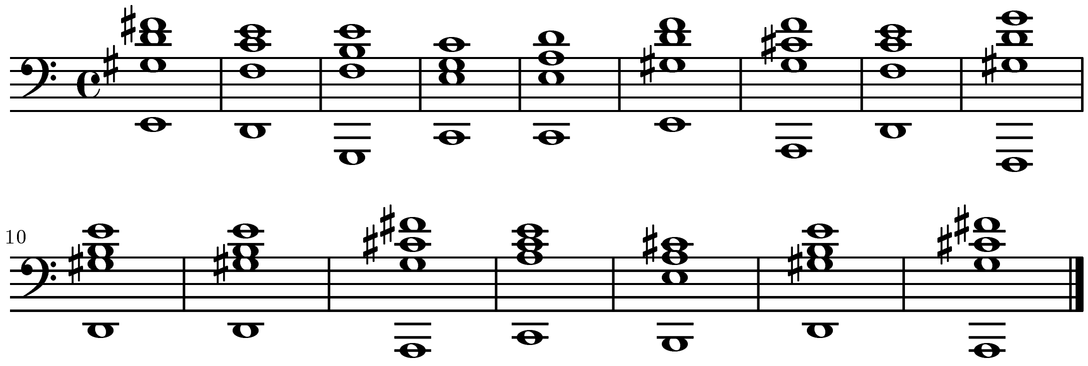

Figure 4. A chord progression generated by the HMM model. While the individual chord transitions are smooth and likely, there is no global chord structure.

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

Eck, D., & Schmidhuber, J. (2002). Finding temporal structure in music: Blues improvisation with LSTM recurrent networks. *Neural Networks for Signal Processing XII, Proc. 2002 IEEE Workshop* (pp. 747–756). New York: IEEE.

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

Kuusi, T. (2001). *Set-class and chord: Examining connection between theoretical ressemblance and perceived closeness*. No. 12 in Studia Musica. Sibelius Academy.

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

Lavrenko, V., & Pickens, J. (2003). Polyphonic music modeling with random fields. *Proceedings of ACM Multimedia*. Berkeley, CA.

Lerdahl, F., & Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. Cambridge, Mass.: MIT Press.

Levine, M. (1990). *The Jazz Piano Book*. Sher Music Co./Advance Music.

Moore, B. (1982). *An Introduction to the Psychology of Hearing*. Academic Press.

Raphael, C., & Stoddard, J. (2003). Harmonic analysis with probabilistic graphical models. *Proceedings of ISMIR 2003*.

Sher, C. (Ed.). (1988). *The New Real Book*, vol. 1. Sher Music Co.

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

Vassilakis, P. (1999). Chords as spectra, harmony as timbre. *J. Acoust. Soc. Am.*, *106*, 2286.