Probabilistic Segmentation of Musical Sequences using Restricted Boltzmann Machines

Stefan Lattner[1], Maarten Grachten[1], Kat Agres[2], Carlos E. Cancino Chacón[1]

[1] Austrian Research Institute for Artificial Intelligence
[2] Queen Mary, University of London

Abstract. A salient characteristic of human perception of music is that musical events are perceived as being grouped temporally into structural units such as phrases or motifs. Segmentation of musical sequences into structural units is a topic of ongoing research, both in cognitive psychology and music information retrieval. Computational models of music segmentation are typically based either on explicit knowledge of music theory or human perception, or on statistical and information-theoretic properties of musical data. The former, rule-based approach has been found to better account for (human annotated) segment boundaries in music than probabilistic approaches [13], although the statistical model proposed in [13] performs almost as well as state-of-the-art rule-based approaches. In this paper, we propose a new probabilistic segmentation method, based on Restricted Boltzmann Machines (RBM). By sampling, we determine a probability distribution over a subset of visible units in the model, conditioned on a configuration of the remaining visible units. We apply this approach to an n-gram representation of melodies, where the RBM generates the conditional probability of a note given its n-1 predecessors. We use this quantity in combination with a threshold to determine the location of segment boundaries. A comparative evaluation shows that this model slightly improves segmentation performance over the model proposed in [13], and as such is closer to the state-of-the-art rule-based models.

1 Introduction

Across perceptual domains, grouping and segmentation mechanisms are crucial for our disambiguation and interpretation of the world. Both top-down, schematic processing mechanisms and bottom-up, grouping mechanisms contribute to our ability to break the world down into meaningful, coherent "chunks". Indeed, a salient characteristic of human perception of music is that musical sequences are not experienced as an indiscriminate stream of events, but rather as a sequence of temporally contiguous musical groups or segments. Elements within a group are perceived to have a coherence that leads to the perception of these events as a structural unit (e.g., a musical phrase or motif).

The origin and nature of this sense of musical coherence, or lack thereof, which gives rise to musical grouping and segmentation has been a topic of ongoing research. A prominent approach from music theory and cognitive psychology has been to apply perceptual grouping mechanisms, such as those suggested by Gestalt psychology, to music perception. *Gestalt principles*, such as the laws of proximity, similarity, and closure, were first discussed in visual perception [20], and have been successfully applied to auditory scene analysis [2] and inspired theories of music perception [11,12,10]. Narmours Implication-Realization theory [12], for example, uses measures of pitch proximity and closure that offer insight into how listeners perceive the boundaries between musical phrases. This type of theory-driven approach has given rise to various rule-based computational models of segmentation. This class of models relies upon the specification of one or more principles according to which musical sequences are grouped.

A second class of computational methods is based on statistical and information theoretic properties of musical data. Recent research in this area has used the statistical structure of sequential tonal and temporal information to compute measures of information (such as Information Content), which serve as a proxy for expectedness (see for example, [1]). Measures of expectation may then be used to calculate segmentation boundaries. For example, a highly expected musical event followed by an unexpected event is often indicative of a perceptual boundary.

A comparison of rule-based and probabilistic approaches [13] suggests the most effective segmentation methods are generally theory-based approaches. The statistical model proposed in [13] (IDyOM) is capable of much better segmentations than simpler statistical models based on digram transition probabilities and point-wise mutual information [3], but still falls slightly short of state-of-the-art rule-based models. Even if rule-based models currently outperform statistical models, there is a motivation to further pursue statistical models of melodic segmentation.

It is plausible that the rules put forth in music-theoretic and perception-based models have been induced by regularities in musical and auditory stimuli. Models that learn directly from the statistics of such stimuli are conceptually simpler than models that describe the perceptual mechanisms of human beings that have internalized the regularities of those stimuli.

In this paper, we introduce a new probabilistic segmentation method, based on a class of stochastic neural networks known as Restricted Boltzmann Machines (RBMs). We present a Monte-Carlo method to determine a probability distribution over a subset of visible units in the model, conditioned on a configuration of the remaining visible units. Processing melodies as n-grams, the RBM generates the conditional probability of a note given its n-1 predecessors. This quantity, in combination with a threshold, determines the location of segment boundaries. In Section 2 we give a brief overview of both rule-based and statistical models for melodic segmentation, where we restrict ourselves to an overview of the models with which we compare our approach: those evaluated in [13]. Then, we will argue that our model (explained in Section 3) has advantages over statistical models based on n-gram counting. In addition to this qualitative comparison of our method to other approaches (Section 3), we reproduce a quantitative evaluation experiment by Pearce et al. [13] (Section 4). The results, as reported and discussed in Section 5, show that our model slightly improves segmentation performance over IDyOM, and as such is closer to the state-of-the-art rule-based models. Finally, we present conclusions and directions for future work in Section 6.

2 Related Work

2.1 Rule based segmentation

One of the first models of melodic segmentation based on Gestalt rules was proposed by Tenney and Polansky in [17]. This theory quantifies rules of local detail to predict grouping judgements. However, this theory does not account for vague or ambiguous grouping judgements, and the selection of their numerical weights is rather arbitrary [17,10]. One of the most popular music theoretic approaches is Lerdahl and Jackendoff's Generative Theory of Tonal Music (GTTM) [10]. This theory pursues the formal description of musical intuitions of experienced listeners through a combination of cognitive principles and generative linguistic theory. In GTTM, the hierarchical segmentation of a musical piece into motifs, phrases and sections is represented through a *grouping structure*. This structure is expressed through consecutively numbered *grouping preference rules* (GPRs), which model possible structural descriptions that correspond to experienced listeners' hearing of a particular piece [10]. According to GTTM, two types of evidence are involved in the determination of the grouping structure. The first kind of evidence to perceive a phrase boundary between two melodic events is *local detail*, i.e. relative temporal proximity like slurs and rests (GPR 2a), inter-onset-interval (IOI) (GPR 2b) and change in register (GPR 3a), dynamics (GPR 3b), articulation (GPR 3c) or duration (GPR 3d).

The organization of *larger-level* grouping involves intensification of the effects picked out by GPRs 2 and 3 on a larger temporal scale (GPR 4), symmetry (GPR 5) and parallelism (GPR 6). While Lerdahl and Jackendoff's work did not attempt to quantify these rules, a computational model for identification of segment boundaries that numerically quantifies the GPRs 2a, 2b, 3a and 3d was proposed by Frankland and Cohen [5]. This model encodes melodic profiles using absolute duration of the notes, and MIDI note numbers for representing absolute pitch.

A related model to the quantification of the GPRs was proposed by Cambouropoulos [4]. The Local Boundary Detection Model (LBDM) consists of a *change* rule and a *proximity* rule, operated over melodic profiles that encode pitch, IOI and rests. On the one hand, the change rule identifies the strength of a segment boundary in relation to the degree of change between consecutive intervals (similar to GPR 3). On the other hand, the proximity rule considers the size of the intervals involved (as in GPR 2). The total boundary strength is then computed as a weighted sum of the boundaries for pitch, IOI and rests, where the weights were empirically selected.

Temperley [16] introduced a similar method, called Grouper, that partitions a melody (represented by onset time, off time, chromatic pitch and a level in a metrical hierarchy) into non-overlapping groups. Grouper uses three *phase structure preference rules* (PSPR) to asses the existence of segment boundaries. PSPR 1 locates boundaries at large IOIs and large offset-to-onset intervals (OOIs), and is similar to GPR 2, while PSPR 3 is a rule for metrical parallelism, analogous to GPR 6. PSPR 2 relates to the length of the phrase, and was empirically determined by Temperley using the Essen Folk Song Collection (EFSC), and therefore, may not be a general rule [14].

2.2 Statistical and information theoretic segmentation

In [14], Pearce, Müllensiefen and Wiggins applied two information theoretic approaches, originally designed by Brent [3] for word identification in unsegmented speech, to construct boundary strength profiles (BSPs) for melodic events. This method relies on the assumption that segmentation boundaries are located in places where certain information theoretic measures have a higher numerical value than in the immediately neighbouring locations. The first approach constructs BSPs using transition probability (TP), the conditional probability of an element of a sequence given the preceding element, while the second method relies on pointwise mutual information (PMI), that measures to which degree the occurrence of an event reduces the model's uncertainty about the co-occurrence of another event, to produce such BSPs

Inspired by developments in musicology, computational linguistics and machine learning, Pearce, Müllensiefen and Wiggins offered the IDyOM model. IDyOM is an unsupervised, multi-layer, variable-order Markov model that computes the conditional probability and Information Content (IC) of a musical event, given the prior context. An overview of IDyOM can be found in [13].

3 Method

The primary assumption underlying statistical models of melodic segmentation is that the perception of segment boundaries is induced by the statistical properties of the data. RBMs (Section 3.2) can be trained effectively as a generative probabilistic model of data (Section 3.5), and are therefore a good basis for defining a segmentation method. However, in contrast to sequential models such as recurrent neural networks, RBMs are models of static data, and do not model temporal dependencies. A common way to deal with this is to feed the model sub-sequences of consecutive events (n-grams) as if they were static entities, without explicitly encoding time. This n-gram approach allows the model to capture regularities among events that take place within an n-gram. With some simplification we can state that these regularities take the form of a joint probability distribution over all events in an n-gram. With Monte-Carlo methods, we can use this joint distribution to approximate the conditional probability of some of these events, given others. This procedure is explained in Sections 3.3 and 3.4.

3.1 Relation to other statistical models

Although our RBM-based method works with n-gram representations just as the statistical methods discussed in Section 2.2, the approaches are fundamentally different. Models such as IDyOM, TP and PMI are based on n-gram counting, and as such has to deal with the trade-off between longer n-grams and sparsity of data that is inevitable when working with longer sub-sequences. In IDyOM, this problem is countered with “back-off” a heuristic to dynamically decrease or increase the n-gram size as the sparsity of the data allows. In contrast, an RBM does not assign probabilities to n-grams based directly on their frequency counts. The non-linear connections between visible units (via a layer of hidden units) allow a much smoother probability distribution, that can also assign non-zero probability to n-grams that were never presented as training data. As a result, it is possible to work with a fixed, relatively large n-gram size, without the need to reduce the size in order to counter data sparsity.

Every computational model requires a set of basic features that describe musical events. In IDyOM, these basic features are treated as statistically independent, and dependencies between features are modelled explicitly by defining combined viewpoints as cross-products of subsets of features. An advantage of the RBM model is that dependencies between features are modelled as an integral part of learning, without the need to specify subsets of features explicitly.

Finally, the statistical methods discussed in Section 2 are fundamentally n-gram based, and it is not obvious how these methods can be adapted to work with polyphonic music rather than monophonic melodies. Although the RBM model presented here uses an n-gram representation, it is straight-forward to adopt the same segmentation approach using a different representation of musical events, such as the note-centred representation proposed in [7]. This would make the RBM suitable for segmenting polyphonic music.

3.2 Restricted Boltzmann Machines

An RBM is a stochastic Neural Network with two layers, a visible layer with units $\mathbf{v} \in \{0,1\}^r$ and a hidden layer with units $\mathbf{h} \in \{0,1\}^q$ [9]. The units of both layers are fully interconnected with weights $\mathbf{W} \in \mathbb{R}^{r \times q}$, while there are no connections between the units within a layer.

In a trained RBM, the marginal probability distribution of a visible configuration $\mathbf{v}$ is given by the equation

$$p(\mathbf{v}) = \frac{1}{Z} \sum_{\mathbf{h}} e^{-E(\mathbf{v},\mathbf{h})}, \tag{1}$$

where $E(\mathbf{v},\mathbf{h})$ is an energy function. The computation of this probability distribution is usually intractable, because it requires summing over all possible joint configurations of $\mathbf{v}$ and $\mathbf{h}$ as

$$Z = \sum_{\mathbf{v},\mathbf{h}} e^{-E(\mathbf{v},\mathbf{h})}. \tag{2}$$

3.3 Approximation of the probability of v

Another way to compute the probability of a visible unit configuration $\mathbf{v}$ is to approximate it through Monte Carlo techniques. To that end, for N randomly initialized *fantasy particles*[1] $\mathbf{Q}$, we execute Gibbs sampling until thermal equilibrium. In the visible *activation vector* $\mathbf{q}_i$ of a fantasy particle i, element q_{ij} specifies the probability that visible unit j is on. Since all visible units are independent given $\mathbf{h}$, the probability of $\mathbf{v}$ based on one fantasy particle's visible activation is computed as:

$$p(\mathbf{v}|\mathbf{q}_i) = \prod_j p(v_j|q_{ij}). \tag{3}$$

As we are using binary units, such an estimate can be calculated by using a binomial distribution with one trial per unit. We average the results over N fantasy particles, leading to an increasingly close approximation of the true probability of $\mathbf{v}$ as N increases:

$$p(\mathbf{v}|\mathbf{Q}) = \frac{1}{N}\sum_i^N \prod_j \binom{1}{v_j} q_{ij}^{v_j}(1-q_{ij})^{1-v_j}. \tag{4}$$

3.4 Posterior probabilities of visible units

When the visible layer consists of many units, N will need to be very large to obtain good probability estimates with the method described above. However, for conditioning a (relatively small) subset of visible units $\mathbf{v}_y \subset \mathbf{v}$ on the remaining visible units $\mathbf{v}_x = \mathbf{v} \setminus \mathbf{v}_y$, the above method is very useful. This can be done by Gibbs sampling after randomly initializing the units $\mathbf{v}_y$ while clamping all other units $\mathbf{v}_x$ according to their initial state in $\mathbf{v}$. In Eq. 4, all $\mathbf{v}_x$ contribute a probability of 1, which results in the conditional probability of $\mathbf{v}_y$ given $\mathbf{v}_x$.

We use this approach to condition the units belonging to the last time step of an n-gram on the units belonging to preceding time steps. For the experiments reported in this paper, we found that it is sufficient to use 150 fantasy particles and for each to perform 150 Gibbs sampling steps.

3.5 Training

We train a single RBM using *persistent contrastive divergence* (PCD) [18] with *fast weights* [19], a variation of the standard *contrastive divergence* (CD) algorithm [8]. PCD is more suitable for sampling than CD, because it results in a better approximation of the likelihood gradient.

Based on properties of neural coding, sparsity and selectivity can be used as constraints for the optimization of the training algorithm [6]. Sparsity encourages competition between hidden units, and selectivity prevents over-dominance

[1] See [18]

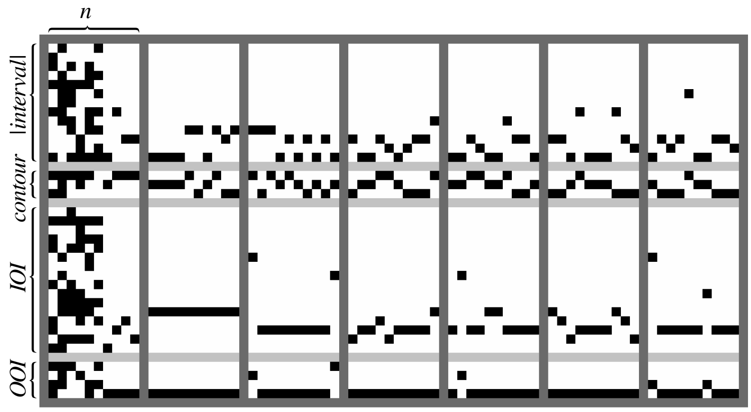

Fig. 1. Seven examples of n-gram training instances (n=10) used as input to the RBM. Within each instance (delimited by a dark gray border), each of the 10 columns represents a note. Each column consists of four *one-hot* encoded viewpoints: |*interval*|, *contour*, *IOI* and *OOI* (indicated by the braces on the left). The viewpoints are separated by horizontal light gray lines for clarity. The first instance shows an example of noise padding (in the first six columns) to indicate the beginning of a melody.

by any individual unit. A parameter μ specifies the desired degree of sparsity and selectivity, whereas another parameter ϕ determines how strongly the sparsity/selectivity constraints are enforced.

3.6 Data Representation

From the monophonic melodies, we construct a set of n-grams by using a sliding window of size n and a step size of 1. For each note in the n-gram, four basic features are computed: 1) absolute values of the pitch interval between the note and its predecessor (in semitones); 2) the contour (up, down, or equal); 3) inter-onset-interval (IOI); and 4) onset-to-offset-interval (OOI). The IOI and OOI values are quantized into semiquaver and quaver, respectively. Each of these four features is represented as a binary vector and its respective value for any note is encoded in a one-hot representation. The first n-1 n-grams in a melody are noise-padded to account for the first n-1 prefixes of the melody. Some examples of binary representations of n-grams are given in Figure 1).

3.7 Information Content

After training the model as described in 3.5, we estimate the probability of the last note conditioned on its preceding notes for each n-gram as introduced in 3.4. From the probabilities $p(e_t \mid e_{t-n+1}^{t-1})$ computed thus, we calculate the IC as:

$$h(e_t \mid e_{t-n+1}^{t-1}) = log_2 \frac{1}{p(e_t \mid e_{t-n+1}^{t-1})}, \tag{5}$$

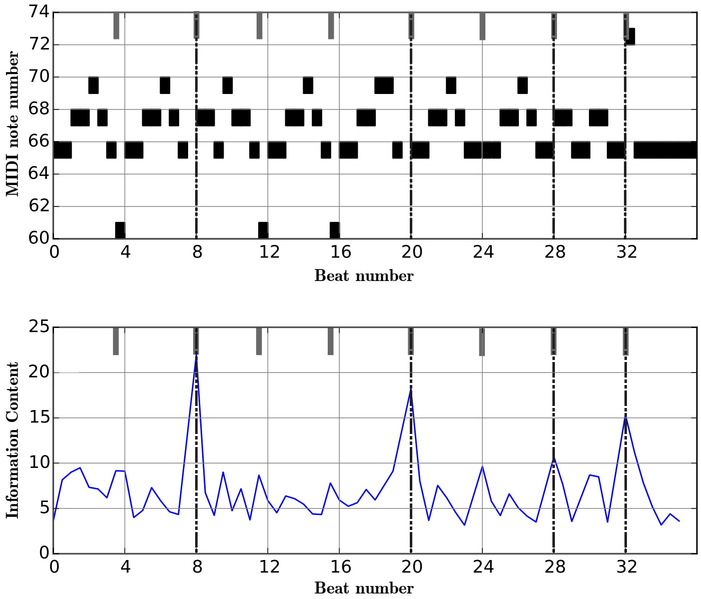

Fig. 2. A BSP calculated from 11-grams. The upper figure shows the notes of 9 measures (36 beats) of a German folk song. The lower figure shows a BSP (i.e. IC) used for segmentation. The correct segmentation (ground truth) is depicted as vertical grey bars at the top of the figures, segment boundaries found by our model are shown as dashed vertical lines. Note that the BSP has particularly high peaks at rests and at high intervals. However, the segment boundary found at beat 28 does not have any of those cues and was still correctly classified.

where e_t is a note event at time step t, and e_k^l is a note sequence from position k to l of a melody. IC is a measure of the unexpectedness of an event given its context. According to a hypothesis of [13], segmentation in auditory perception is determined by perceptual expectations for auditory events. In this sense, the IC relates directly to this perceived boundary strength, thus we call the IC over a note sequence *boundary strength profile*.

3.8 Peak Picking

Based on the BSP described in the previous section, we need to find a concrete binary segmentation vector. For that, we use the peak picking method described in [13]. This method finds all peaks in the profile and keeps those which are k times the standard deviation greater than the mean boundary strength, linearly weighted from the beginning of the melody to the preceding value:

$$S_n > k\sqrt{\frac{\sum_{i=1}^{n-1}(w_i S_i - \bar{S}_{w,1\ldots n-1})^2}{\sum_1^{n-1} w_i}} + \frac{\sum_{i=1}^{n-1} w_i S_i}{\sum_1^{n-1} w_i}, \tag{6}$$

where S_m is the m-th value of the BSP, and w_i are the weights which emphasize recent values over those of the beginning of the song (triangular window), and k has to be found empirically.

4 Experiment

4.1 Training Data

In this work, we use the EFSC [15]. This database is a widely used corpus in MIR for experiments on symbolic music. This collection consists of more than 6000 transcriptions of folk songs primarily from Germany and other European regions. The EFSC collection is commonly used for testing computational models of music segmentation, due to the fact that it is annotated with phrase markers.

In accordance with [13], we used the *Erk* subset of the EFSC, which consists of 1705 German folk melodies with a total of $78,995$ note events. Phrase boundary annotations are marked at about 12% of the note events.

4.2 Procedure

The model is trained and tested on the data described in Section 4.1, with n-gram lengths varying between 1 and 11. For each n-gram length, we perform 5-fold cross-validation and average the results over all folds. Similar to the approach in [13], after computing the BSPs, we evaluate different k from the set $\{0.70, 0.75, 0.80, 0.85, 0.90, 0.95, 1.00\}$ and choose the value that maximizes F1 for the respective n-gram length. To make results comparable to those reported in [13], the output of the model is appended with an implicit (and correct) phrase boundary at the end of each melody.

Since the hyper-parameters of the model are inter-dependent, it is infeasible to exhaustively search for the optimal parameter setting. For the current experiment, we have manually chosen a set of hyper-parameters that give reasonable results for the different models tested: 200 hidden units, a batch size of 100, a momentum of 0.6, and a learning rate of 0.007 which we linearly decrease to zero during training. The fast weights used in the training algorithm (see Section 3.5) help the fantasy particles mix well, even with small learning rates. The learning rate of the fast weights is increased from 0.002 to 0.007 during training. The training is continued until convergence of the parameters (typically between 100 and 300 epochs). The sparsity parameters (see Section 3.5) are set to $\mu = 0.04$, and $\phi = 0.65$, respectively. In addition, we use a value of 0.0035 for $L2$ weight regularization, which penalizes large weight coefficients.

5 Results and Discussion

We tested three different representations for pitch, yielding the following F1 scores for 10-grams: *absolute pitch* (0.582), *interval* (0.600), and the absolute value of interval (i.e. $|$*interval*$|$) plus *contour* (0.602). The latter representation

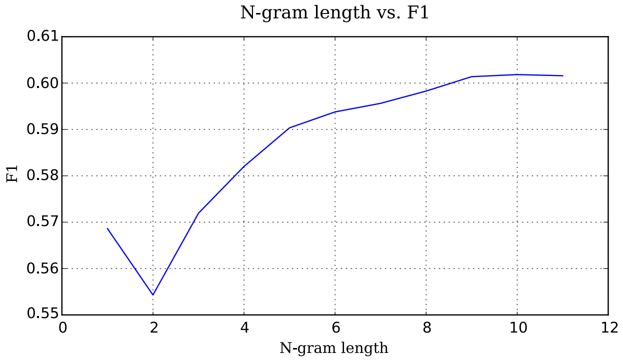

Fig. 3. Maximal F1 scores for different n-gram lengths.

was chosen for our experiments, as it showed the best performance. Not surprisingly, relative pitch representations lead to better results, as they reduce the number of combination possibilities in the input. Event though the difference in F1 score between *interval* and |*interval*| *plus contour* representation is not significant, it still shows that it is valid to decompose viewpoints into their elementary informative parts. Such an approach, next to reducing the input dimensionality, may also support the generalization ability of a model (e.g. |interval| representation in music may help to understand the concept of inversion).

Figure 3 shows the F1 score obtained by models of different n-gram sizes. The fact that boundary detection is reasonably good even for 1-grams is likely due to the fact that the 1-gram includes the OOI, which is mostly zero, except for the relatively rare occurrence of a rest between notes. Because of this, the probability values assigned to OOI values by a trained RBM behave like an inverted rest indicator: high for OOI = 0, and low for OOI > 0. This makes the behaviour of the 1-gram RBM much like that of the GPR 2a rule (Section 2.1). That GPR 2a performs slightly better (see Table 1) can be explained by the fact that the RBM also detects segment boundaries at large (and unlikely) pitch intervals, which are not always correct.

Another remarkable result is that 1-grams perform better than 2-grams (see Figure 3). Although we have no definite explanation for this yet, the difference may be related to the fact that in the 1-gram model, the probability is estimated by sampling without clamping any units. In contrast, for 2-grams half the units get clamped during sampling. Prior tests with our method for computing the conditional probability (Section 3.4) have revealed (unsurprisingly) that the quality of the approximation decreases with the ratio of unknown units over given (clamped) units. This phenomenon may also partly account for the steady increase of F1 scores for increasing n-grams sizes larger than one. Nevertheless, the increasing performance with increasing n-gram size demonstrates that the RBM based segmentation method is less susceptible to problems of data sparseness encountered in n-gram counting approaches.

Model	Precision	Recall	F1
Grouper	0.71	0.62	0.66
LBDM	0.70	0.60	0.63
RBM (10-gram)	0.83	0.50	0.60
IDyOM	0.76	0.50	0.58
GPR 2a	0.99	0.45	0.58
GPR 2b	0.47	0.42	0.39
GPR 3a	0.29	0.46	0.35
GPR 3d	0.66	0.22	0.31
PMI	0.16	0.32	0.21
TP	0.17	0.19	0.17
Always	0.13	1.00	0.22
Never	0.00	0.00	0.00

Table 1. Results of the model comparison, ordered by F1 score. Table adapted from [13], with permission.

6 Conclusion

In this paper, an RBM-based unsupervised probabilistic method for segmentation of melodic sequences was presented. In contrast to other statistical methods, our method does not rely on frequency counting, and thereby circumvents problems related to data sparsity. The method performs slightly better than IDyOM, a sophisticated frequency counting model.

The segment boundary detection capabilities of our model are still slightly lower than state-of-the-art rule based methods that rely on gestalt principles formulated for musical stimuli. This result underlines the remaining challenge to find segmentation models that correspond to human perception, based only on musical stimuli in combination with universal learning principles.

An important aspect of human perception that is missing in our current method is equivalent of short-term memory, to bias long-term expectations based on the stimuli in the direct past (see [13]). Furthermore, we wish to investigate the effect of different architectural factors on the segmentation behaviour of the model, like an increased number of hidden layers, or an increased number of hidden units per hidden layer. Lastly, the formation of boundary strength profiles may be improved by involving other information theoretic quantities, such as the entropy of conditional probability distributions.

Acknowledgements

The project Lrn2Cre8 acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET grant number 610859. We thank Marcus Pearce for sharing the Essen data used in [13].

References

1. Agres, K., Abdallah, S., Pearce, M.: An information-theoretic account of musical expectation and memory. In: Proc. of the 35th Annual Conference of the Cognitive Science Society. pp. 127–132. Cognitive Science Society, Austin, Texas (2013)
2. Bregman, A.S.: Auditory Scene Analysis. MIT Press, Cambridge, MA (1990)
3. Brent, M.R.: An efficient, probabilistically sound algorithm for segmentation and word discovery 34(1–3), 71–105 (1999)
4. Cambouropoulos, E.: The local boundary detection model (LBDM) and its application in the study of expressive timing. In: Proceedings of the International Computer Music Conference. pp. 17–22. San Francisco (2001)
5. Frankland, B.W., Cohen, A.J.: Parsing of Melody: Quantification and Testing of the Local Grouping Rules of Lerdahl and Jackendoff's A Generative Theory of Tonal Music. Music Perception 21(4), 499–543 (2004)
6. Goh, H., Thome, N., Cord, M.: Biasing restricted Boltzmann machines to manipulate latent selectivity and sparsity. NIPS workshop on deep learning and unsupervised feature learning (2010)
7. Grachten, M., Krebs, F.: An assessment of learned score features for modeling expressive dynamics in music. IEEE Transactions on Multimedia 16(5), 1211–1218 (2014), `http://dx.doi.org/10.1109/TMM.2014.2311013`
8. Hinton, G.E., Osindero, S., Teh, Y.: A fast learning algorithm for deep belief nets. Neural Computation 18, 1527–1554 (2006)
9. Hinton, G.E.: Training products of experts by minimizing contrastive divergence. Neural Computation 14(8), 1771–1800 (Jul 2002)
10. Lerdahl, F., Jackendoff, R.: A generative theory of tonal music. MIT press (1983)
11. Meyer, L.: Emotion and meaning in Music. University of Chicago Press, Chicago (1956)
12. Narmour, E.: The analysis and cognition of basic melodic structures : the Implication-Realization model. University of Chicago Press (1990)
13. Pearce, M., Müllensiefen, D., Wiggins, G.A.: Melodic Grouping in Music Information Retrieval: New Methods and Applications. Advances in Music Information Retrieval 274(Chapter 16), 364–388 (2010)
14. Pearce, M.T., Müllensiefen, D., Wiggins, G.: The role of expectation and probabilistic learning in auditory boundary perception: A model comparison. Perception 39(10), 1367–1391 (2010)
15. Schaffrath, H.: The Essen Folksong Collection in Kern Format. In: Huron, D. (ed.) Database containing , folksong transcriptions in the Kern format and a -page research guide computer database. Menlo Park, CA (1995)
16. Temperley, D.: The Cognition of Basic Musical Structures. MIT Press, Cambridge, MA (2001)
17. Tenney, J., Polansky, L.: Temporal Gestalt Perception in Music. Journal of Music Theory 24(2), 205–241 (1980)
18. Tieleman, T.: Training Restricted Boltzmann Machines using Approximations to the Likelihood Gradient. In: Proceedings of the 25th international conference on Machine learning. pp. 1064–1071. ACM New York, NY, USA (2008)
19. Tieleman, T., Hinton, G.: Using Fast Weights to Improve Persistent Contrastive Divergence. In: Proceedings of the 26th international conference on Machine learning. pp. 1033–1040. ACM New York, NY, USA (2009)
20. Wertheimer, M.: Laws of organization in perceptual forms. A source book of Gestalt psychology pp. 71–88 (1938)