uses of the model possible.

The hierarchical approach presented in this paper fits well with the hierarchical structure of music in frequency as well as in time domains. Each part of the model represents an explainable entity (e.g. tone partial, pitch, chord). In contrast to the DBNs, each part of the model can be visualized. Visualization not only exposes the layered structure of the model, but also discloses information processed by the observed part and its influence on other parts and their activations. This insight into the music signal can be used in several scenarios — music visualization, music analysis and music synthesis.

We have developed a real-time visualization of the model, enabling deeper understanding of the processed information. When observing an inferred audio signal, the output of all layers of the model is presented by visualizing activations of parts. This insight enables detailed analysis of each event in the music signal and may bring additional event details to light. For example, a chord inversion can be observed by looking into the activated subtree of the chord from top layers to bottom-ones. Thus, visualization of our model offers an innovative user interface for music analysis.

The transparency of the model can also be exploited for music processing and synthesis. Parts across all layers form a variety of harmonic structures, and can be used for signal manipulation and synthesis. By activating a set of parts at different locations, a new spectral representation is produced. Although the interface may not provide a sufficient amount of features for a standalone music performance, it can be used as a sound generator in a combination with a music instrument, e.g. a MIDI keyboard. The interface thus serves as an advanced tool for spectral modification, while the instrument provides the interface for performance.

6. CONCLUSION AND FUTURE WORK

This paper presents a compositional hierarchical model as an alternative to deep learning architectures based on neural networks. The model shares a great deal of similarities with other deep architectures, including a multi-layer structure, unsupervised generative learning and suitability for discriminative tasks. Furthermore, the white-box structure of the model offers new utilizations of the model. We highlighted three possible applications: feature extraction for MIR tasks, music visualization and music analysis/synthesis.

The model's internals rely on findings in the fields of neurobiology and cognitive sciences. By implementing biologically-inspired mechanisms into the model, we made an attempt to build a model which partially resembles a subset of functions of the human auditory system. We intend to retain and further develop this aspect of the model with an intention to bring the computational modeling closer to human auditory perception.

The paper presents an initial development of our model. We plan to further extend it with the focus on temporal modeling. Parts can namely be extended into the time domain, thus bringing their activations closer to event-based modeling. We also plan to tackle temporal tasks, such as onset detection, as well as beat tracking and tempo estimation. The proposed model is also going to be evaluated for pattern analysis of symbolic data, including discovery of repeated themes, and symbolic melodic similarity.

7. REFERENCES

[1] Eric Battenberg and David Wessel. Analyzing Drum Patterns using Conditional Deep Belief Networks. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 37–42, 2012.

[2] Juan P. Bello and Jeremy Pickens. A robust mid-level representation for harmonic content in music signals. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 304–311, London, 2005.

[3] J. Stephen Downie, Andreas F. Ehmann, Mert Bay, and M. Cameron Jones. The Music Information Retrieval Evaluation eXchange: Some Observations and Insights. In Wieczorkowska A.A. and Ras Z.W., editors, *Advances in Music Information Retrieval*, pages 93–115. Springer-Verlag, Berlin, 2010.

[4] Valentin Emiya, Roland Badeau, and Bertrand David. Multipitch Estimation of Piano Sounds Using a New Probabilistic Spectral Smoothness Principle. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(6):1643–1654, August 2010.

[5] Philippe Hamel and Douglas Eck. Learning Features from Music Audio with Deep Belief Networks. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 339–344, 2010.

[6] Eric J. Humphrey, Juan P. Bello, and Yann LeCun. Moving beyond feature design: deep architectures and automatic feature learning in music informatics. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Porto, 2012.

[7] Honglak Lee, Peter Pham, Yan Largman, and Andrew Y. Ng. Unsupervised feature learning for audio classification using convolutional deep belief networks. In *Advances in Neural Information Processing Systems*, pages 1096–1104, 2009.

[8] Aleš Leonardis and Sanja Fidler. Towards scalable representations of object categories: Learning a hierarchy of parts. *Computer Vision and Pattern Recognition, IEEE*, pages 1–8, 2007.

[9] Abdel-rahman Mohamed, George E. Dahl, and Geoffrey Hinton. Acoustic Modeling using Deep Belief Networks. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(1):14–22, 2010.

[10] Nicola Orio. Music Retrieval: A Tutorial and Review. *Foundations and Trends® in Information Retrieval*, 1(1):1–90, 2006.

[11] Helene Papadopoulos and Geoffroy Peeters. Large-case Study of Chord Estimation Algorithms Based on Chroma Representation and HMM. *Content-Based Multimedia Indexing*, 53-60, 2007.

[12] Eric M. Schmidt and Youngmoo E. Kim. Learning Rhythm and Melody Features with Deep Belief Networks. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 21–26, 2013.

[13] Erik M. Schmidt and Youngmoo E. Kim. Learning emotion-based acoustic features with deep belief networks. In *2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 65–68. IEEE, October 2011.

[14] Felix Weninger, Christian Kirst, Bjorn Schuller, and Hans-Joachim Bungartz. A discriminative approach to polyphonic piano note transcription using supervised non-negative matrix factorization. In *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 6–10, Vancouver, 2013.

[15] Dong Yu and Li Deng. Deep Learning and Its Applications to Signal and Information Processing [Exploratory DSP. *IEEE Signal Processing Magazine*, 28(1):145–154, January 2011.