# References – IEEE Volume 36 January 2019 Automatic Music Transcription

[1] A. Klapuri and M. Davy, Eds., Signal Processing Methods for Music Transcription. New York: Springer-Verlag, 2006.

* General overview of field
* Errors present in current models
* Characteristics of musical notes
* Description of unpitched sound problem Ch. 5

[2] E. Benetos, S. Dixon, D. Giannoulis, H. Kirchhoff, and A. Klapuri, “Automatic music transcription: Challenges and future directions,” J. Intelligent Inform. Syst., vol. 41, no. 3, pp. 407–434, 2013.

* General overview of field
* Future challenges
* Polyphonic instruments challenge
* Problems with current models

[3] M. Müller, D. P. Ellis, A. Klapuri, and G. Richard, “Signal processing for music analysis,” IEEE J. Sel. Topics Signal Process., vol. 5, no. 6, pp. 1088–1110, 2011.

* Relation to source separation

[4] M. Schedl, E. Gómez, and J. Urbano, “Music information retrieval: Recent developments and applications,” Foundations Trends Inform. Retrieval, vol. 8, pp. 127–261, 2014. doi: 10.1561/1500000042.

* Relation to Information Retrieval tasks like music similarity

[5] N. Boulanger-Lewandowski, Y. Bengio, and P. Vincent, “Modeling temporal dependencies in high-dimensional sequences: Application to polyphonic music generation and transcription,” in Proc. Int. Conf. Machine Learning, 2012, pp. 1159– 1166.

* Relation to Information Retrieval tasks like music similarity
* Neural Network (NN) and Music Language Model (MLM)
* NN MNE does not consider note interactions
* NN MNE considers note interactions
* Polyphonic music detection via RBM machine with RNN to postprocess the acoustic output of an AMT system
* Modelling high dimensional sequences with deep learning methods

[6] T. Virtanen, M. D. Plumbley, and D. P. W. Ellis, Eds., Computational Analysis of Sound Scenes and Events. New York: Springer-Verlag, 2018.

* AMT and sound event detection (SED)
* Similarity of methodologies used in SED with AMT methods

[7] L. Su and Y.-H. Yang, “Escaping from the abyss of manual annotation: New methodology of building polyphonic datasets for automatic music transcription,” in Proc. Int. Symp. Computer Music Multidisciplinary Research, 2015, pp. 309–321.

* Way to circumvent lack of ground truth transcriptions using music performers
* Lack of annotated datasets

[8] Z. Duan, B. Pardo, and C. Zhang, “Multiple fundamental frequency estimation by modeling spectral peaks and non-peak regions,” IEEE Trans. Audio, Speech, Language Process. (2006–2013), vol. 18, no. 8, pp. 2121–2133, 2010.

* Probabilistic MPE modelling

[9] Z. Duan and D. Temperley, “Note-level music transcription by maximum likelihood sampling,” in Proc. 15th Int. Society Music Information Retrieval Conf., 2014, pp. 181–186.

* Spectral likelihood model for MNE considers note interactions

[10] Z. Duan, J. Han, and B. Pardo, “Multi-pitch streaming of harmonic sound mixtures,” IEEE/ACM Trans. Audio, Speech, Language Process., vol. 22, no. 1, pp. 138–150, 2014.

* MPS instrument tracking

[11] V. Emiya, R. Badeau, and B. David, “Multipitch estimation of piano sounds using a new probabilistic spectral smoothness principle,” IEEE Trans. Audio, Speech, Language Process. (2006–2013), vol. 18, no. 6, pp. 1643–1654, 2010.

* MPE using traditional Signal Processing techniques
* Largest dataset of several hours of piano music
* Challenges to NN

[12] L. Su and Y.-H. Yang, “Combining spectral and temporal representations for multipitch estimation of polyphonic music,” IEEE/ACM Trans. Audio, Speech, Language Process., vol. 23, no. 10, pp. 1600–1612, 2015.

* MPE using traditional Signal Processing techniques
* MNE with median filtering does not consider note interactions

[13] P. H. Peeling, A. T. Cemgil, and S. J. Godsill, “Generative spectrogram factorization models for polyphonic piano transcription,” IEEE Trans. Audio, Speech, Language Process. (2006–2013), vol. 18, no. 3, pp. 519–527, 2010.

* Bayesian MPE

[14] P. Smaragdis and J. C. Brown, “Non-negative matrix factorization for polyphonic music transcription,” in Proc. IEEE Workshop Applications Signal Processing Audio and Acoustics, 2003, pp. 177–180.

* NMF MPE

[15] E. Vincent, N. Bertin, and R. Badeau, “Adaptive harmonic spectral decomposition for multiple pitch estimation,” IEEE Trans. Audio, Speech, Language Process. (2006–2013), vol. 18, no. 3, pp. 528–537, 2010.

* NMF MPE
* Supervised NMF narrow-band sub templates to enforce harmonic structure for all NMF templates
* Remove interference from templates by extracting templates from recordings with only one note

[16] E. Benetos and S. Dixon, “Multiple-instrument polyphonic music transcription using a temporally-constrained shift-invariant model,” J. Acoust. Soc. Amer., vol. 133, no. 3, pp. 1727–1741, 2013.

* NMF MPE
* MPS instrument tracking
* Shift invariant dictionaries can represent multiple F0s
* Sharing parameters between pitches
* Markov process to model note evolutionary phases

[17] B. Fuentes, R. Badeau, and G. Richard, “Harmonic adaptive latent component analysis of audio and application to music transcription,” IEEE Trans. Audio, Speech, Language Process. (2006–2013), vol. 21, no. 9, pp. 1854–1866, 2013.

* NMF MPE
* Shift invariant dictionaries can represent multiple F0s
* Sharing parameters between pitches

[18] S. Sigtia, E. Benetos, and S. Dixon, “An end-to-end neural network for polyphonic piano music transcription,” IEEE/ACM Trans. Audio, Speech, Language Process., vol. 24, no. 5, pp. 927–939, 2016.

* NN MPE
* MLM considers note interactions
* Long range dependencies between notes using an acoustic front end with a symbolic level module resembling an MLM
* Use of MIDI data a recurrent NN was trained to predict active notes in the next frame given those in the past
* NADE architecture and WaveNet
* RNN-RBM MM combined with acoustic predictions using a probabilistic graph model

[19] R. Kelz, M. Dorfer, F. Korzeniowski, S. Böck, A. Arzt, and G. Widmer, “On the potential of simple framewise approaches to piano transcription,” in Proc. Int. Society Music Information Retrieval Conf., 2016, pp. 475–481.

* NN MPE
* Large scale hyperparameter search describing individual system components
* Performed well against other models

[20] J. Nam, J. Ngiam, H. Lee, and M. Slaney, “A classification-based polyphonic piano transcription approach using learned feature representations,” in Proc. Int. Society Music Information Retrieval Conf., 2011, pp. 175–180.

* HMM MNE does not consider note interactions

[21] M. Marolt, “A connectionist approach to automatic transcription of polyphonic piano music,” IEEE Trans. Multimedia, vol. 6, no. 3, pp. 439–449, 2004.

* MNE interonset interval approach considers note interactions
* Marolts Sonic System based on NNs
* Analyzing the output of adaptive oscillators to track and group partials in the output of a gammatone filterbank

[22] A. Cogliati, Z. Duan, and B. Wohlberg, “Context-dependent piano music transcription with convolutional sparse coding,” IEEE/ACM Trans. Audio, Speech, Language Process., vol. 24, no. 12, pp. 2218–2230, 2016.

* MNE considers note interactions through unifed framework that estimates pitch,onset and offset together
* Context specific transcription : single instruments have become much more accurate

[23] S. Ewert and M. B. Sandler, “Piano transcription in the studio using an extensible alternating directions framework,” IEEE/ACM Trans. Audio, Speech, Language Process., vol. 24, no. 11, pp. 1983–1997, 2016.

* MNE considers note interactions through unifed framework that estimates pitch,onset and offset together
* Markov process to model note evolutionary phases
* Comparison to Marolt’s Sonic System
* Limitations of NNs: cold start problem
* Context specific transcription : single instruments have become much more accurate

[24] C. Hawthorne, E. Elsen, J. Song, A. Roberts, I. S. C. Raffel, J. Engel, S. Oore, and D. Eck, “Onsets and frames: Dual-objective piano transcription,” in Proc. Int. Society Music Information Retrieval Conf., 2018, pp. 50–57.

* MNE considers note interactions through unifed framework that estimates pitch,onset and offset together
* Google Brain general purpose piano transcription
* Two networks: one focuses on note onsets and informs the network focused on perceiving note lengths
* Limitations of NNs: cold start problem

[25] V. Arora and L. Behera, “Multiple F0 estimation and source clustering of polyphonic music audio using PLCA and HMRFs,” IEEE/ACM Trans. Audio, Speech, Language Processing, vol. 23, no. 2, pp. 278–287, 2015.

* MPS instrument tracking

[26] E. Cambouropoulos, “Pitch spelling: A computational model,” Music Perception, vol. 20, no. 4, pp. 411–429, 2003.

* Notation level pitch Spelling

[27] H. Grohganz, M. Clausen, and M. Mueller, “Estimating musical time information from performed MIDI files,” in Proc. Int. Society Music Information Retrieval Conf., 2014, pp. 35–40.

* Notation level timing quantization

[28] I. Karydis, A. Nanopoulos, A. Papadopoulos, E. Cambouropoulos, and Y. Manolopoulos, “Horizontal and vertical integration/segregation in auditory streaming: A voice separation algorithm for symbolic musical data,” in Proc. Sound and Music Computing Conf., 2007, pp. 299–306.

* Notation level voice separation

[29] A. Cogliati, D. Temperley, and Z. Duan, “Transcribing human piano performances into music notation,” in Proc. Int. Society Music Information Retrieval Conf., 2016, pp. 758–764.

* Notation level method to convert MIDI to music notation with comparison to commercial packages and performance transcription

[30] R. G. C. Carvalho and P. Smaragdis, “Towards end-to-end polyphonic music transcription: Transforming music audio directly to a score,” in 2017 IEEE Workshop Applications Signal Processing Audio and Acoustics, 2017, pp. 151–155.

* End-to-end proof of concept work audio-to-music notation

[31] D. D. Lee and H. S. Seung, “Algorithms for non-negative matrix factorization,” in Proc. Advances Neural Information Processing Systems, 2001, pp. 556–562.

* Deriving update rules for descent-based minimization of divergence in NMF between V and DA (spectro temporal representation and matrix product of the Dictionary and Activation matrices
* Explanation and advantages of NMF algorithms

[32] S. A. Abdallah and M. D. Plumbley, “Unsupervised analysis of polyphonic music by sparse coding,” IEEE Trans. Neural Netw. (1990–2011), vol. 17, no. 1, pp. 179–196, 2006.

* Challenges of NMF including correlation of sound objects and inharmonicity
* Sparse coding approach to obtain solutions that are dominated by activations that are few but substantial
* Unsupervised NMF

[33] I. Goodfellow, A. Courville, and Y. Bengio, Deep Learning. Cambridge, MA: MIT Press, 2016.

* NNs learning nonlinear functions and optimization methods
* Stochastic gradient descent
* Convolutional networks in the time direction
* Long Short-Term Memory Layers
* Interpolation of templates and valid spectra manifolds in NN

[34] S. Böck and M. Schedl, “Polyphonic piano note transcription with recurrent neural networks,” in Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, 2012, pp. 121–124.

* NNs revival
* Two spectrograms for high time and frequency resolution
* Benefits of LTSMs to model long range dependencies
* Probability of measuring a particular chord sequence

[35] S. Wang, S. Ewert, and S. Dixon, “Identifying missing and extra notes in piano recordings using score-informed dictionary learning,” IEEE/ACM Trans. Audio, Speech, Language Process., vol. 25, no. 10, pp. 1877–1889, 2017.

* Score informed transcription

[36] X. Serra, “A multicultural approach in music information research,” in Proc. 12th Int. Society Music Information Retrieval Conf., 2011, pp. 151–156.

* Non-Western music problem