

5. CONCLUSIONS

We proposed a new unsupervised Bayesian nonnegative matrix factorization approach to extract the singing voice from background music accompaniment and illustrated the novelty on an analytical and true optimum solution to the Poisson-Exponential BNMF. Through the VB-EM inference procedure, the proposed method automatically selected different number of bases to fit various experimental conditions. We conducted two clustering algorithms to find the grouping of bases into vocal and music sources. Experimental results showed the consistent improvement of using BNMF factorization with NMF clustering over the other singing-voice separation methods in terms of GNSDR. In future works, the proposed BNMF shall be extended to multi-layer source separation and applied to detect unknown number of sources.

6. REFERENCES

- [1] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer Science, 2006.
- [2] N. Boulanger-Lewandowski, G. J. Mysore, and M. Hoffman. Exploiting long-term temporal dependencies in NMF using recurrent neural networks with application to source separation. In *Proc. of ICASSP*, pages 337–344, 2014.
- [3] A. S. Bregman. *Auditory Scene Analysis: the Perceptual Organization of Sound*. MIT Press, 1990.
- [4] A. T. Cemgil. Bayesian inference for nonnegative matrix factorisation models. *Computational Intelligence and Neuroscience*, (Article ID 785152), 2009.
- [5] 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 (B)*, 39(1):1–38, 1977.
- [6] H. Fujihara, M. Goto, J. Ogata, and H. G. Okuno. Lyricsynchronizer: automatic synchronization system between musical audio signals and lyrics. *IEEE Journal of Selected Topics in Signal Processing*, 5(6):1252–1261, 2011.
- [7] P. O. Hoyer. Non-negative matrix factorization with sparseness constraints. *The Journal of Machine Learning Research*, 5:1457–1469, 2004.
- [8] C.-L. Hsu and J.-S. R. Jang. On the improvement of singing voice separation for monaural recordings using the MIR-1K dataset. *IEEE Transactions on Audio, Speech, Language Processing*, 18(2):310–319, 2010.
- [9] P.-S. Huang, S. D. Chen, P. Smaragdis, and M. Hasegawa-Johnson. Singing-voice separation from monaural recordings using robust principal component analysis. In *Proc. of ICASSP*, pages 57–60, 2012.
- [10] D. D. Lee and H. S. Seung. Algorithms for non-negative matrix factorization. *Advances in Neural Information Processing Systems*, pages 556–562, 2000.
- [11] A. Mesaros, T. Virtanen, and A. Klapuri. Singer identification in polyphonic music using vocal separation and pattern recognition methods. In *Proc. of Annual Conference of International Society for Music Information Retrieval*, pages 375–378, 2007.
- [12] Z. Rafii and B. Pardo. A simple music/voice separation method based on the extraction of the repeating musical structure. In *Proc. of ICASSP*, pages 221–224, 2011.
- [13] Z. Rafii and B. Pardo. Repeating pattern extraction technique (REPET): A simple method for music/voice separation. *IEEE Transactions on Audio, Speech, Language Processing*, 21(1):73–84, Jan. 2013.
- [14] M. N. Schmidt and M. Morup. Non-negative matrix factor 2-D deconvolution for blind single channel source separation. In *Proc. of ICA*, pages 700–707, 2006.
- [15] M. N. Schmidt, O. Winther, and L. K. Hansen. Bayesian non-negative matrix factorization. In *Proc. of ICA*, pages 540–547, 2009.
- [16] P. Smaragdis. Convolutional speech bases and their application to speech separation. *IEEE Transactions on Audio, Speech, Language Processing*, 15(1):1–12, 2007.
- [17] M. Spiertz and V. Gnann. Source-Filter based clustering for monaural blind source separation. In *Proc. of International Conference on Digital Audio Effects*, pages 1–4, 2009.
- [18] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B*, 58(1):267–288, 1996.
- [19] S. Vembu and S. Baumann. Separation of vocals from polyphonic audio recordings. In *Proc. of ISMIR*, pages 375–378, 2005.
- [20] E. Vincent, R. Gribonval, and C. Fevotte. Performance measurement in blind audio source separation. *IEEE Transaction on Audio, Speech and Language Processing*, 14(4):1462–1469, 2006.
- [21] D. Yang and W. Lee. Disambiguating music emotion using software agents. In *Proc. of ISMIR*, pages 52–57, 2004.
- [22] Y.-H. Yang. On sparse and low-rank matrix decomposition for singing voice separation. In *Proc. of ACM International Conference on Multimedia*, pages 757–760, 2012.
- [23] B. Zhu, W. Li, R. Li, and X. Xue. Multi-stage non-negative matrix factorization for monaural singing voice separation. *IEEE Transactions on Audio, Speech, Language Processing*, 21(10):2096–2107, 2013.