



github.com/alexanderlerch/2024-MIG

music informatics group

overview

alexander lerch

introduction

music informatics group

■ mission

- create new technologies transforming and improving how we *make, produce, perform, discover, and consume music*
- advance the field of AI for audio through *knowledge-driven machine learning*

■ objectives

- enable/improve *machine understanding of music* and musical language
- create *interpretable and controllable systems*
- design algorithms with *low data requirements*



tasks

selected tasks of interest

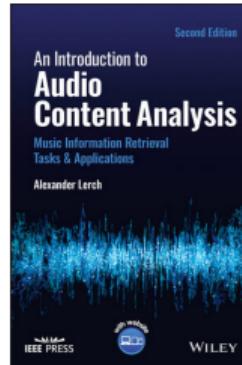
■ audio content analysis [1]

- music/audio *classification*
 - ▶ genre/events [2], [3]
 - ▶ instruments [4], [5], [6]
 - ▶ tagging [6], [7], [8]
 - ▶ emotion [9]
 - ▶ pedestrians [10], [11]
- music *transcription*
 - ▶ drum transcription [12]
 - ▶ chord detection [13]
- music *performance analysis* [14]

■ music source separation [15], [16], [17]

■ sound and music generation

- controllability [18]
- evaluation [19], [20], [21], [22]



tasks

selected tasks of interest

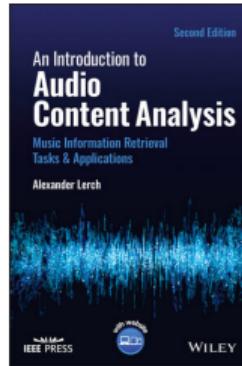
■ audio content analysis [1]

- music/audio *classification*
 - ▶ genre/events [2], [3]
 - ▶ instruments [4], [5], [6]
 - ▶ tagging [6], [7], [8]
 - ▶ emotion [9]
 - ▶ pedestrians [10], [11]
- music *transcription*
 - ▶ drum transcription [12]
 - ▶ chord detection [13]
- music *performance analysis* [14]

■ music source separation [15], [16], [17]

■ sound and music generation

- controllability [18]
- evaluation [19], [20], [21], [22]



tasks

selected tasks of interest

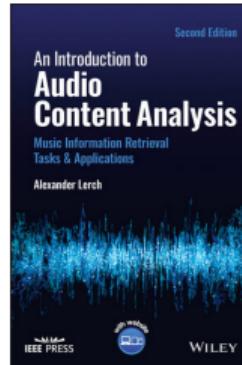
■ audio content analysis [1]

- music/audio *classification*
 - ▶ genre/events [2], [3]
 - ▶ instruments [4], [5], [6]
 - ▶ tagging [6], [7], [8]
 - ▶ emotion [9]
 - ▶ pedestrians [10], [11]
- music *transcription*
 - ▶ drum transcription [12]
 - ▶ chord detection [13]
- music *performance analysis* [14]

■ music source separation [15], [16], [17]

■ sound and music generation

- controllability [18]
- evaluation [19], [20], [21], [22]



methods

methods of interest

■ knowledge transfer & representation learning

- improved structure of embedded representations [23], [24]
- enforcing the meaning of specific embedding dimensions [18], [20]
- knowledge transfer [6], [7], [26]

■ low-resource machine learning

- semi- and self-supervised learning [4], [25]
- reprogramming [3], [5]



■ objective evaluation of generative systems

- evaluation of controllable systems with correlated attributes [21], [27]
- statistical models for comparison of properties [19]
- metrics for sound generation [22]

methods

methods of interest

■ knowledge transfer & representation learning

- improved structure of embedded representations [23], [24]
- enforcing the meaning of specific embedding dimensions [18], [20]
- knowledge transfer [6], [7], [26]

■ low-resource machine learning

- semi- and self-supervised learning [4], [25]
- reprogramming [3], [5]



■ objective evaluation of generative systems

- evaluation of controllable systems with correlated attributes [21], [27]
- statistical models for comparison of properties [19]
- metrics for sound generation [22]

methods

methods of interest

■ knowledge transfer & representation learning

- improved structure of embedded representations [23], [24]
- enforcing the meaning of specific embedding dimensions [18], [20]
- knowledge transfer [6], [7], [26]

■ low-resource machine learning

- semi- and self-supervised learning [4], [25]
- reprogramming [3], [5]



■ objective evaluation of generative systems

- evaluation of controllable systems with correlated attributes [21], [27]
- statistical models for comparison of properties [19]
- metrics for sound generation [22]

links

links

music informatics group: musicinformatics.gatech.edu

book: www.AudioContentAnalysis.org

alexander lerch (alexander.lerch@gatech.edu):

- www.linkedin.com/in/lerch
- scholar.google.com/citations?user=29dF3UIAAAAJ
- github.com/alexanderlerch



references |

references

- [1] A. Lerch, *An Introduction to Audio Content Analysis: Music Information Retrieval Tasks and Applications*, en-us, 2nd ed. Hoboken, N.J.: Wiley-IEEE Press, 2023, ISBN: 978-1-119-89094-2. Accessed: Nov. 4, 2022. [Online]. Available: <https://ieeexplore.ieee.org/servlet/opac?bknumber=9965970>.
- [2] J. J. Burred and A. Lerch, "Hierarchical Automatic Audio Signal Classification," *Journal of the Audio Engineering Society (JAES)*, vol. 52, no. 7/8, pp. 724–739, 2004. [Online]. Available: http://www.musicinformatics.gatech.edu/wp-content_nondefault/uploads/2016/10/Burred-and-Lerch-2004-Hierarchical-Automatic-Audio-Signal-Classification.pdf.
- [3] Y.-N. Hung, C.-H. H. Yang, P.-Y. Chen, and A. Lerch, "Low-Resource Music Genre Classification with Cross-Modal Neural Model Reprogramming," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Rhodes Island, Greece: Institute of Electrical and Electronics Engineers (IEEE), 2023. DOI: [10.1109/ICASSP49357.2023.10096568](https://doi.org/10.1109/ICASSP49357.2023.10096568). [Online]. Available: <https://arxiv.org/abs/2211.01317>.
- [4] S. Gururani and A. Lerch, "Semi-Supervised Audio Classification with Partially Labeled Data," in *Proceedings of the IEEE International Symposium on Multimedia (ISM)*, online: Institute of Electrical and Electronics Engineers (IEEE), 2021. [Online]. Available: <https://arxiv.org/abs/2111.12761>.
- [5] H.-H. Chen and A. Lerch, "Music Instrument Classification Reprogrammed," in *Proceedings of the International Conference on Multimedia Modeling (MMM)*, Bergen, Norway, 2023. [Online]. Available: <https://arxiv.org/abs/2211.08379>.
- [6] Y. Ding and A. Lerch, "Audio Embeddings as Teachers for Music Classification," in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Milan, Italy, 2023. DOI: [10.48550/arXiv.2306.17424](https://doi.org/10.48550/arXiv.2306.17424). Accessed: Jul. 3, 2023. [Online]. Available: <http://arxiv.org/abs/2306.17424>.

references II

references

- [7] Y. Ding and A. Lerch, "Embedding Compression for Teacher-to-Student Knowledge Transfer," in *Proceedings of the International Conference on Acoustics Speech and Signal Processing (ICASSP) - Satellite Workshop Deep Neural Network Model Compression*, Seoul, Korea: Institute of Electrical and Electronics Engineers (IEEE), Feb. 2024. DOI: [10.48550/arXiv.2402.06761](https://doi.org/10.48550/arXiv.2402.06761). Accessed: Feb. 27, 2024. [Online]. Available: <http://arxiv.org/abs/2402.06761>.
- [8] T. A. Ma and A. Lerch, "Music auto-tagging in the long tail: A few-shot approach," in *Proceedings of the AES Convention*, New York, Sep. 2024. DOI: [10.48550/arXiv.2409.07730](https://doi.org/10.48550/arXiv.2409.07730). Accessed: Sep. 13, 2024. [Online]. Available: <http://arxiv.org/abs/2409.07730>.
- [9] K. N. Watcharasupat, Y. Ding, T. A. Ma, P. Seshadri, and A. Lerch, "Uncertainty Estimation in the Real World: A Study on Music Emotion Recognition," in *Proceedings of the European Conference on Information Retrieval (ECIR)*, Lucca, Italy: arXiv, 2025. DOI: [10.48550/arXiv.2501.11570](https://doi.org/10.48550/arXiv.2501.11570). Accessed: Jan. 30, 2025. [Online]. Available: <http://arxiv.org/abs/2501.11570>.
- [10] P. Seshadri, C. Han, B.-W. Koo, N. Posner, S. Guhathakurta, and A. Lerch, "ASPED: An Audio Dataset for Detecting Pedestrians," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Seoul: Institute of Electrical and Electronics Engineers (IEEE), 2024. DOI: [10.48550/arXiv.2309.06531](https://doi.org/10.48550/arXiv.2309.06531). Accessed: Dec. 14, 2023. [Online]. Available: <http://arxiv.org/abs/2309.06531>.
- [11] C. Han et al., "Understanding Pedestrian Movement Using Urban Sensing Technologies: The Promise of Audio-based Sensors," en, *Urban Informatics*, vol. 3, no. 1, p. 22, Jul. 2024, ISSN: 2731-6963. DOI: [10.1007/s44212-024-00053-9](https://doi.org/10.1007/s44212-024-00053-9). Accessed: Jul. 10, 2024. [Online]. Available: <https://doi.org/10.1007/s44212-024-00053-9>.
- [12] C.-W. Wu et al., "A Review of Automatic Drum Transcription," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 9, pp. 1457–1483, 2018, ISSN: 2329-9290. DOI: [10.1109/TASLP.2018.2830113](https://doi.org/10.1109/TASLP.2018.2830113). [Online]. Available: http://www.musicinformatics.gatech.edu/wp-content_nondefault/uploads/2018/05/Wu-et-al.-2018-A-review-of-automatic-drum-transcription.pdf.

references III

references

- [13] X. Zhou and A. Lerch, "Chord Detection Using Deep Learning," in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Malaga: ISMIR, 2015. DOI: [10.5281/zenodo.1416968](https://doi.org/10.5281/zenodo.1416968). [Online]. Available: http://www.musicinformatics.gatech.edu/wp-content_nondefault/uploads/2015/10/Zhou_Lerch_2015_Chord-Detection-Using-Deep-Learning.pdf.
- [14] K. A. Pati, S. Gururani, and A. Lerch, "Assessment of Student Music Performances Using Deep Neural Networks," en, *Applied Sciences*, vol. 8, no. 4, p. 507, 2018. DOI: [10.3390/app8040507](https://doi.org/10.3390/app8040507). Accessed: Mar. 27, 2018. [Online]. Available: <http://www.mdpi.com/2076-3417/8/4/507/pdf>.
- [15] Y.-N. Hung and A. Lerch, "Multi-Task Learning for Instrument Activation Aware Music Source Separation," in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Montreal: International Society for Music Information Retrieval (ISMIR), 2020. [Online]. Available: https://musicinformatics.gatech.edu/wp-content_nondefault/uploads/2020/08/Hung-and-Lerch-2020-Multi-Task-Learning-for-Instrument-Activation-Awar.pdf.
- [16] K. N. Watcharasupat et al., "A Generalized Bandsplit Neural Network for Cinematic Audio Source Separation," *IEEE Open Journal of Signal Processing*, vol. 5, pp. 73–81, 2024, ISSN: 2644-1322. DOI: [10.1109/OJSP.2023.3339428](https://doi.org/10.1109/OJSP.2023.3339428). Accessed: Jan. 2, 2024. [Online]. Available: <https://ieeexplore.ieee.org/document/10342812>.
- [17] K. N. Watcharasupat and A. Lerch, "A Stem-Agnostic Single-Decoder System for Music Source Separation Beyond Four Stems," in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, San Francisco, Jun. 2024. DOI: [10.48550/arXiv.2406.18747](https://doi.org/10.48550/arXiv.2406.18747). Accessed: Aug. 8, 2024. [Online]. Available: <http://arxiv.org/abs/2406.18747>.
- [18] K. A. Pati and A. Lerch, "Attribute-based Regularization for Latent Spaces of Variational Auto-Encoders," *Neural Computing and Applications*, 2020. DOI: [10.1007/s00521-020-05270-2](https://doi.org/10.1007/s00521-020-05270-2). [Online]. Available: <https://arxiv.org/pdf/2004.05485.pdf>.

references IV

references

- [19] L.-C. Yang and A. Lerch, "On the Evaluation of Generative Models in Music," en, *Neural Computing and Applications*, no. 32, pp. 4773–4784, 2020, ISSN: 1433-3058. DOI: [10.1007/s00521-018-3849-7](https://doi.org/10.1007/s00521-018-3849-7). Accessed: Nov. 4, 2018. [Online]. Available: http://www.musicinformatics.gatech.edu/wp-content_nondefault/uploads/2018/11/postprint.pdf.
- [20] A. Pati and A. Lerch, "Is Disentanglement Enough? On Latent Representations for Controllable Music Generation," en, in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Online, 2021, p. 8. [Online]. Available: <https://arxiv.org/abs/2108.01450>.
- [21] K. N. Watcharasupat, J. Lee, and A. Lerch, "Latte: Cross-framework Python Package for Evaluation of Latent-based Generative Models," en, *Software Impacts*, p. 100222, 2022, ISSN: 26659638. DOI: [10.1016/j.simpa.2022.100222](https://doi.org/10.1016/j.simpa.2022.100222). Accessed: Jan. 13, 2022. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S2665963822000033>.
- [22] A. Vinay and A. Lerch, "Evaluating Generative Audio Systems and their Metrics," in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, arXiv:2209.00130 [cs, eess], Bangalore, IN, Aug. 2022. DOI: [10.48550/arXiv.2209.00130](https://doi.org/10.48550/arXiv.2209.00130). Accessed: Sep. 3, 2022. [Online]. Available: <http://arxiv.org/abs/2209.00130>.
- [23] P. Seshadri and A. Lerch, "Improving Music Performance Assessment with Contrastive Learning," en, in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Online, 2021, p. 8. [Online]. Available: <https://arxiv.org/abs/2108.01711>.
- [24] A. B. Ma and A. Lerch, "Representation Learning for the Automatic Indexing of Sound Effects Libraries," in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, arXiv:2208.09096 [cs, eess], Bangalore, IN, Aug. 2022. DOI: [10.48550/arXiv.2208.09096](https://doi.org/10.48550/arXiv.2208.09096). Accessed: Aug. 22, 2022. [Online]. Available: <http://arxiv.org/abs/2208.09096>.
- [25] C.-W. Wu and A. Lerch, "From Labeled to Unlabeled Data – On the Data Challenge in Automatic Drum Transcription," in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Paris, 2018. [Online]. Available: http://www.musicinformatics.gatech.edu/wp-content_nondefault/uploads/2018/06/Wu-and-Lerch-From-Labeled-to-Unlabeled-Data-On-the-Data-Chal.pdf.

references V

references

- [26] Y.-N. Hung and A. Lerch, "Feature-informed Embedding Space Regularization for Audio Classification," in *Proceedings of the European Signal Processing Conference (EUSIPCO)*, Belgrade, Serbia, 2022. doi: [10.48550/arXiv.2206.04850](https://doi.org/10.48550/arXiv.2206.04850). [Online]. Available: <http://arxiv.org/abs/2206.04850>.
- [27] K. N. Watcharasupat and A. Lerch, "Evaluation of Latent Space Disentanglement in the Presence of Interdependent Attributes," in *Late Breaking Demo (Extended Abstract), Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Online, 2021. Accessed: Nov. 11, 2021. [Online]. Available: <http://arxiv.org/abs/2110.05587>.