Learning with Music Signals: Technology Meets Education

Meinard Müller

International Audio Laboratories Erlangen[†], Erlangen, Germany

Abstract

Music information retrieval (MIR) is an exciting and challenging research area that aims to develop techniques and tools for organizing, analyzing, retrieving, and presenting music-related data. Being at the intersection of engineering and humanities, MIR relates to different research disciplines, including signal processing, machine learning, information retrieval, musicology, and the digital humanities. In this tutorial, using music as a tangible and concrete application domain, we approach the concept of learning from different angles, addressing technological and educational aspects. In this way, the tutorial serves several purposes: it gives a gentle introduction to MIR while introducing a new software package for teaching and learning music processing, it highlights avenues for developing explainable machine-learning models, and it discusses how recent technology can be applied and communicated in interdisciplinary research and education.

CCS Concepts

• ***Information systems*** → ***Music retrieval;*** • ***Applied computing*** → ***Sound and music computing;***

1. Music Information Retrieval

Music is a ubiquitous and vital part of our lives. Thanks to the proliferation of digital music services, we have access to music nearly anytime and anywhere, and we interact with music in a variety of ways, both as listeners and active participants. As a result, music has become one of the most popular categories of multimedia content. In general terms, music processing research aims to contribute concepts, models, and algorithms that extend our capabilities of accessing, analyzing, understanding, and creating music [MPMV19]. In particular, the development of computational tools that allow users to find, organize, analyze, and interact with music has become central to the research field known as Music Information Retrieval (MIR). Given the complexity and diversity of music, research has to account for various aspects such as the genre, the instrumentation, the musical form, melodic and harmonic properties, dynamics, tempo, rhythm, and timbre, to name a few. Music signals typically comprise a wide range and a large number of different sound sources. Postprocessing and the use of audio effects in the mixing and mastering stages may further complicate the analysis of recorded musical material. Furthermore, music is inherently multimodal, incorporating speech-like signals (singing), video (of live performances), and static images (scanned music scores), see Figure 1. This wealth of data makes MIR an interdisciplinary and challenging field of research, which closely connects to technical disciplines such as signal processing, machine learning, and information retrieval as well as other disciplines such as mathematics, musicology, and the digital humanities, see Figure 2.

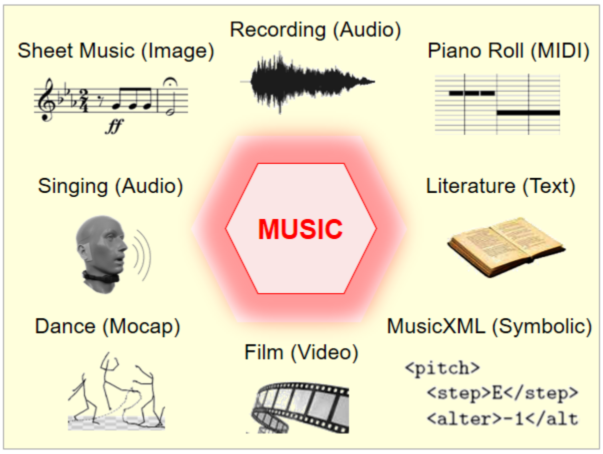

Figure 1: *Music as a multimodal and challenging application domain.*

2. General Goal of Tutorial

As an overarching goal of this tutorial, we want to show how music can serve as a challenging and instructive multimedia domain to break new ground in technology and education in various disci-

[†] The International Audio Laboratories Erlangen are a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and Fraunhofer Institut für Integrierte Schaltungen IIS.

DOI: 10.2312/egt.20231032

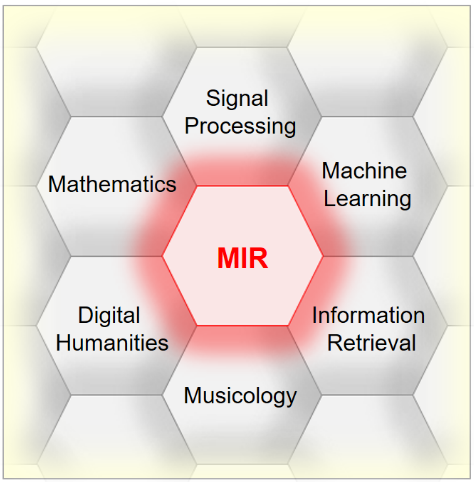

Figure 2: *Music Information Retrieval (MIR) and its relationships to other research disciplines.*

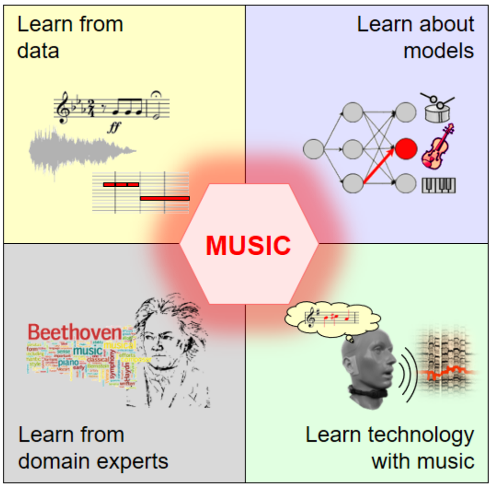

Figure 3: *Holistic approach to learning with music signals.*

plines, as shown in Figure 2. In the context of concrete and tangible MIR applications, we approach and explore the concept of learning from different angles, see Figure 3. First, learning from data, we discuss recent deep learning (DL) techniques for extracting complex features and hidden relationships directly from raw music signals. Second, by learning from the experience of traditional engineering approaches, we indicate how to better understand existing and develop more interpretable DL-based systems by integrating prior knowledge in various ways. In particular, we show how one may transform classical model-based MIR approaches into differentiable multilayer networks, which can then be blended with DL-based techniques to form explainable hybrid models that are less vulnerable to data biases and confounding factors. Third, we give examples of collaborations with domain experts, considering specialized music corpora to gain a deeper understanding of both the music data and the models' behavior while exploring the potential of computational models for musicological research. Fourth, we discuss how music may serve as a motivating vehicle to make learning in technical disciplines such as signal processing or machine learning an interactive pursuit.

In summary, the tutorial's novelty lies in presenting a holistic approach to learning using music as a challenging and tangible application domain. In this way, the tutorial serves several purposes: it gives a gentle introduction to MIR while introducing a new software package for teaching and learning music processing, it highlights avenues for developing explainable machine-learning models, and it discusses how recent technology can be communicated in interdisciplinary research and education.

3. Further Discussion and Literature

In the following, we give more details about the tutorial's technological and educational aspects, assuming the different perspectives of learning as indicated by Figure 3. Furthermore, we provide some references to the literature for further studies.

3.1. Machine Learning for Music Signal Processing

As in general multimedia processing, many of the recent advances in MIR and music signal processing have been driven by techniques based on deep learning (DL). For example, DL-based techniques have led to significant improvements for many MIR tasks such as music transcription [BDDE19, CC19, USW*19, WDS*18], chord recognition [KW16b, POGS19, WL19], local key estimation [WSM20], melody estimation [BMS*17, DEP19], beat tracking [BKW16], tempo estimation [BDK19, SM18], and lyric alignment [SED18], to name a few. In particular, significant improvements could be achieved for specific music scenarios where sufficient training data is available. For example, deep learning has revolutionized the research area of source separation thanks to the availability of suitable training data in the form of multitrack music recordings (e.g., having separated tracks for vocals, accompaniment, drums, and bass) [RLS*18, SED18, SLI18, SULM19, HKVM20, ÖM22]. An excellent tutorial and open source tools for music source separation are provided by [MSS20]. As deep learning advances, one can observe a paradigm shift from knowledge-driven feature engineering to data-driven feature learning [GBC16, HBL12]. For example, in the music context, DL-based techniques have been increasingly used to learn and adapt feature representations for harmonic analysis from training examples [KW16a, WZZ*21]. These are just a few examples of how deep learning has found its way into the MIR area. Thanks to their ability to learn (rather than hand-design) features as part of a model, DL-based techniques seem to increasingly dominate previous engineering techniques. This trend has been reinforced by the availability of (open-source) software and hardware developments and the increasing availability of digitized data.

3.2. Interpretable Models and Knowledge Integration

On the downside, DL-based approaches also come at a cost, being a data-hungry and computing-intensive technology. Furthermore, the behavior of DL-based systems is often hard to understand, and the trained models may capture information that is not directly related to the core problem. These general properties of DL-based approaches can also be observed when analyzing and processing music, which spans an enormous range of forms and styles—not to speak of the many ways music may be generated and represented. While in music analysis and classification problems, one aims at capturing musically relevant aspects related to melody, harmony, rhythm, or instrumentation, data-driven approaches often capture confounding factors that may not directly relate to the target concept. One main advantage of classical model-based engineering approaches is that they result in explainable and explicit models that can be adjusted in intuitive ways. On the downside, such hand-engineered approaches not only require profound signal processing skills and domain knowledge but may also result in highly specialized solutions that cannot be directly transferred to other problems.

In the report [MBN*22], one finds a detailed discussion on the benefits and limitations of recent deep learning techniques using music as a challenging application domain. One general strategy to obtain explainable models that are less vulnerable to data biases and confounding factors is to combine deep learning with classical model-based strategies by integrating knowledge at various stages of a DL-based processing pipeline. For example, one may exploit knowledge already at the input level by using data representations that better isolate information known to be relevant to a task and remove information known to be irrelevant. Knowledge can also be exploited in the design of the output representation (e.g., structured output spaces for chord recognition that account for bass, root, and chroma [MB17]) or the loss function used for optimization (e.g., considering hierarchical consistency measures [WCB18, KM22]). Furthermore, during the data generation and training process, one may use musically informed data augmentation techniques to enforce certain invariances [MHB15, SG15, THFK18]. As another general strategy, one may incorporate musical domain knowledge via the model architecture to better use its capacity to account for particular acoustic or structural aspects. For example, many classical engineering pipelines (e.g., for beat tracking or chord recognition) consist of a series of convolution, rectification, and pooling operations, which can be reinterpreted and implemented as a multilayer network with specific weights and activation functions. In recent years, there has been a research trend in building hybrid neural networks with fixed components (that mimic traditional engineering approaches) and free components (that develop the full power of deep learning). An example of such an approach is the contribution [EHGR20], which integrates differentiable signal processing modules into deep learning pipelines.

3.3. Music Understanding and Applications

It is a generally accepted fact in machine learning that “better data leads to better models” and that understanding your data (and the task at hand) is essential for advancing DL-based research. Therefore, MIR experts should collaborate with music experts to gain a deeper understanding of the datasets involved while exploring the potential of computational tools within concrete application scenarios of musicological relevance. On a broader level, such interdisciplinary collaborations explore to which extent and how musicological research may benefit from using (recent) computer-based methods [GHGM13, HvKKV17, RSM23, Tza14, VWK11]. On the one hand, MIR research has contributed with technology that allows music scientists to browse, search and analyze extensive music collections concerning musically relevant structures in an intuitive and interactive way. On the other hand, collaborating with music experts introduces new perspectives and scientific challenges in engineering and data science.

As argued before, the success of deep learning depends very much on the availability of (suitably annotated and structured) data. As in other multimedia domains, DL-based approaches in MIR that have been trained on datasets of limited diversity (e.g., using only Western popular music) often result in models that overfit to specific musical and acoustic characteristics. For example, training singing voice separation models mainly on popular music may result in models that perform poorly on opera recordings. Similarly, multi-pitch estimation models trained on Western music that is based on the 12-tone equal-tempered scale may fail for non-Western music based on different tonal concepts. Having cross-disciplinary collaborations with music experts and working with non-standard datasets is essential to gain a deeper understanding of machine learning methods and their (overfitting) behavior, see [WSM20] for an example.

At this point, we want to emphasize that aspects of collecting, accessing, representing, generating, annotating, preprocessing, and structuring music-related data are by far not trivial [PBD18, Ser14]. First of all, music offers a wide range of data types and formats including text, symbolic data, audio, image, and video, see also Figure 1. Then, depending on the MIR task, one may need to deal with various types of annotations, including lyrics, chords, guitar tabs, tapping (beat, measure) positions, album covers, as well as a variety of user-generated tags and other types of metadata. To algorithmically exploit the wealth of these various types of information, one requires methods for linking semantically related data sources (e.g., songs and lyrics, sheet music and recorded performances, lead sheet and guitar tabs) [Mül21b]. Finally, as for data accessibility, copyright issues are a main obstacle for the distribution and usage of music collections in academic research.

3.4. Interactive Learning in Engineering Through Music

In recent years, deep learning and artificial intelligence have become prominent parts of the public narrative. Discussions of techniques, the data they are trained on, and the implications of their use are all of broad relevance to a general audience. Music processing as an interdisciplinary research field may serve as a vehicle to make learning and understanding current trends in machine learning an interactive pursuit. As we discussed before, the inclusion of music bridges the gap between the humanities and other communities, such as mathematics, computer science, and engineering. Therefore, research in MIR may help foster communication and discussions across disciplines on the benefits and limitations of current deep learning technologies.

As argued in the article [MMK21], music may also yield an intuitive entry point to support education on various levels. For example, most students and schoolchildren are familiar with music video games (e.g., SingStar or Rock Band), where the task is to sing along

Proceedings published by Eurographics - The European Association for Computer Graphics.

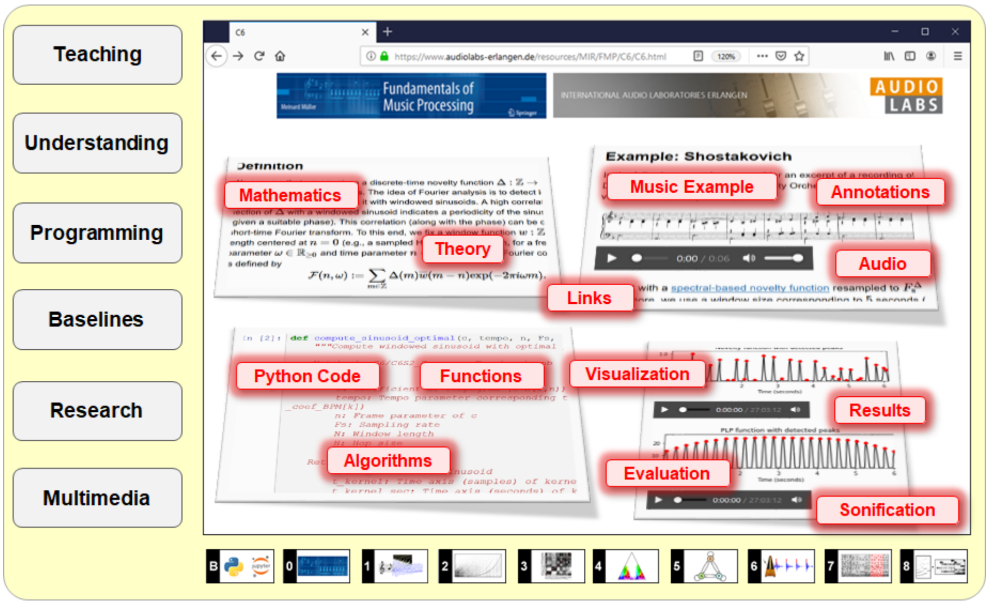

Figure 4: *An overview of educational aspects of the FMP notebooks for teaching and leaning fundamentals of music processing and their implementation using the interactive Jupyter notebook framework. See also* `https://www.audiolabs-erlangen.de/FMP` *and [Mül21a, MZ19].*

with music in order to score points. Similarly, when listening to a piece of music, we are often able to tap along with the musical beat—sometimes, we even do this unconsciously. When teaching a challenging concept, instructors attempt to find compelling examples that their students can hold onto through the sea of equations and subtleties. For technical disciplines such as signal processing and machine learning, MIR tasks such as melody estimation or beat tracking can be those motivating examples and help anchor abstract concepts in a concrete, familiar context. This kind of contextualized pedagogical practice has been shown to improve student retention in computer science curricula [Guz10].

In addition to motivating and tangible music-based scenarios, the availability of suitably designed software packages that make signal processing and machine learning more accessible are crucial in view of interactive learning [Guz13]. Over the last 20 years, as MIR developed as a research field, so did computational accessibility, and the MIR community has contributed with several excellent toolboxes that provide modular source code for processing and analyzing music signals. Prominent examples are `essentia` [BWG*13], `madmom` [BKS*16], `Marsyas` [Tza09], and the `MIRtoolbox` [LT07]. While most of these toolboxes are mainly designed for research-oriented access to audio processing, the Python package `librosa` [MRL*15] with its educational lens has also been incorporated into introductory MIR courses. Specifically designed for educational purposes, the FMP notebooks [Mül21a, MZ19] offer an interactive foundation for teaching and learning fundamentals of music processing (FMP), see Figure 4. The FMP notebooks leverage the web-based Jupyter notebook framework [KRKP*16], which allows users to create documents that contain live code, text-based information, mathematical formulas, plots, images, sound examples, and videos. By closely following the topics of the textbook titled *Fundamentals of Music Processing* [Mül21b], the FMP notebooks provide an explicit link between structured educational environments and current professional practices, inline with current curricular recommendations for computer science [Joi13].

Acknowledgments

I would like to express my gratitude to my previous and current Ph.D. students. This tutorial is based on results and insights achieved in close collaboration with my students and colleagues. Furthermore, I thank the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) for its continuous support over the last decade, which allowed me to conduct fundamental research in music processing. In particular, I would like to thank the DFG for the recently approved project on “Learning with Music Signals” (MU 2686/15-1; Grant No. 500643750), which allows me and my students to continue this strand of research. Finally, I thank my colleagues at the International Audio Laboratories Erlangen—a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and Fraunhofer Institut für Integrierte Schaltungen IIS. Being in the vicinity of both the university and the Audio & Multimedia division of Fraunhofer IIS, the AudioLabs offer an excellent infrastructure that enables close scientific collaborations in an ideal setting.

References

[BDDE19] Benetos E., Dixon S., Duan Z., Ewert S.: Automatic music transcription: An overview. *IEEE Signal Processing Magazine 36*, 1 (2019), 20–30. `doi:10.1109/MSP.2018.2869928`. 2

[BDK19] Böck S., Davies M. E. P., Knees P.: Multi-task learning of tempo and beat: Learning one to improve the other. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)* (Delft, The Netherlands, 2019), pp. 486–493. 2

[BKS*16] Böck S., Korzeniowski F., Schlüter J., Krebs F., Widmer G.: madmom: A new Python audio and music signal processing library. In *Proceedings of the ACM International Conference on Multimedia (ACM-MM)* (Amsterdam, The Netherlands, 2016), pp. 1174–1178. doi:10.1145/2964284.2973795. 4

[BKW16] Böck S., Krebs F., Widmer G.: Joint beat and downbeat tracking with recurrent neural networks. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)* (New York City, New York, USA, 2016), pp. 255–261. doi:10.5281/zenodo.1415835. 2

[BMS*17] Bittner R. M., McFee B., Salamon J., Li P., Bello J. P.: Deep salience representations for F0 tracking in polyphonic music. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)* (Suzhou, China, 2017), pp. 63–70. doi:10.5281/zenodo.1417937. 2

[BWG*13] Bogdanov D., Wack N., Gómez E., Gulati S., Herrera P., Mayor O., Roma G., Salamon J., Zapata J. R., Serra X.: Essentia: An audio analysis library for music information retrieval. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)* (Curitiba, Brazil, 2013), pp. 493–498. doi:10.5281/zenodo.1415016. 4

[CC19] Choi K., Cho K.: Deep unsupervised drum transcription. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)* (Delft, The Netherlands, 2019), pp. 183–191. doi:10.5281/zenodo.3527773. 2

[DEP19] Doras G., Esling P., Peeters G.: On the use of U-net for dominant melody estimation in polyphonic music. In *Proceedings of the International Workshop on Multilayer Music Representation and Processing (MMRP)* (Milan, Italy, 2019), pp. 66–70. 2

[EHGR20] Engel J., Hantrakul L., Gu C., Roberts A.: DDSP: Differentiable digital signal processing. In *Proceedings of the International Conference on Learning Representations (ICLR)* (Virtual, 2020). URL: https://openreview.net/forum?id=B1x1ma4tDr. 3

[GBC16] Goodfellow I., Bengio Y., Courville A.: *Deep Learning*. MIT Press, Cambridge and London, 2016. URL: http://www.deeplearningbook.org. 2

[GHGM13] Gómez E., Herrera P., Gómez-Martin F.: Computational ethnomusicology: Perspectives and challenges. *Journal of New Music Research 42*, 2 (2013), 111–112. doi:10.1080/09298215.2013.818038. 3

[Guz10] Guzdial M.: Does contextualized computing education help? *ACM Inroads 1*, 4 (2010), 4–6. 4

[Guz13] Guzdial M.: Exploring hypotheses about media computation. In *International Computing Education Research Conference (ICER)* (La Jolla, CA, USA, 2013), pp. 19–26. URL: https://doi.org/10.1145/2493394.2493397, doi:10.1145/2493394.2493397. 4

[HBL12] Humphrey E. J., Bello J. P., LeCun Y.: Moving beyond feature design: Deep architectures and automatic feature learning in music informatics. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)* (Porto, Portugal, 2012), pp. 403–408. doi:10.5281/zenodo.1415726. 2

[HKVM20] Hennequin R., Khlif A., Voituret F., Moussallam M.: Spleeter: a fast and efficient music source separation tool with pretrained models. *Journal of Open Source Software 5*, 50 (2020), 2154. Deezer Research. URL: https://doi.org/10.21105/joss.02154, doi:10.21105/joss.02154. 2

[HvKKV17] Holzapfel A., van Kranenburg P., Kursell J., Volk A.: Computational ethnomusicology: Methodologies for a new field. Workhop of the Lorentz Center, Leiden, The Netherlands, https://www.lorentzcenter.nl/lc/web/2017/866/info.php3?wsid=866, 2017. 3

[Joi13] Joint Task Force on Computing Curricula of the Association for Computing Machinery (ACM) and IEEE Computer Society: *Computer Science Curricula 2013: Curriculum Guidelines for Undergraduate Degree Programs in Computer Science*. Association for Computing Machinery, New York, NY, USA, 2013. 4

[KM22] Krause M., Müller M.: Hierarchical classification for singing activity, gender, and type in complex music recordings. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (Virtual and Singapore, 2022), pp. 406–410. doi:10.1109/ICASSP43922.2022.9747690. 3

[KRKP*16] Kluyver T., Ragan-Kelley B., Pérez F., Granger B., Bussonnier M., Frederic J., Kelley K., Hamrick J., Grout J., Corlay S., Ivanov P., Avila D., Abdalla S., Willing C., development team J.: Jupyter notebooks—a publishing format for reproducible computational workflows. In *Proceedings of the International Conference on Electronic Publishing* (Göttingen, Germany, 2016), pp. 87–90. 4

[KW16a] Korzeniowski F., Widmer G.: Feature learning for chord recognition: The deep chroma extractor. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)* (New York City, New York, USA, 2016), pp. 37–43. doi:10.5281/zenodo.1416314. 2

[KW16b] Korzeniowski F., Widmer G.: A fully convolutional deep auditory model for musical chord recognition. In *Proceedings of the IEEE International Workshop on Machine Learning for Signal Processing (MLSP)* (Salerno, Italy, 2016). doi:10.1109/MLSP.2016.7738895. 2

[LT07] Lartillot O., Toiviainen P.: MIR in MATLAB (II): A toolbox for musical feature extraction from audio. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)* (Vienna, Austria, 2007), pp. 127–130. doi:10.5281/zenodo.1417145. 4

[MB17] McFee B., Bello J. P.: Structured training for large-vocabulary chord recognition. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)* (Suzhou, China, 2017), pp. 188–194. doi:10.5281/zenodo.1414880. 3

[MBN*22] Müller M., Bittner R., Nam J., Krause M., Özer Y.: Deep learning and knowledge integration for music audio analysis (Dagstuhl Seminar 22082). *Dagstuhl Reports 12*, 2 (2022), 103–133. URL: https://drops.dagstuhl.de/opus/volltexte/2022/16933, doi:10.4230/DagRep.12.2.103. 3

[MHB15] McFee B., Humphrey E. J., Bello J. P.: A software framework for musical data augmentation. In *Proceedings of International Society for Music Information Retrieval Conference (ISMIR)* (Málaga, Spain, 2015), pp. 248–254. 3

[MMK21] Müller M., McFee B., Kinnaird K.: Interactive learning of signal processing through music: Making Fourier analysis concrete for students. *IEEE Signal Processing Magazine 38*, 3 (2021), 73–84. doi:10.1109/MSP.2021.3052181. 3

[MPMV19] Müller M., Pardo B., Mysore G. J., Välimäki V.: Recent advances in music signal processing. *IEEE Signal Processing Magazine 36*, 1 (2019), 17–19. doi:10.1109/MSP.2018.2876190. 1

[MRL*15] McFee B., Raffel C., Liang D., Ellis D. P., McVicar M., Battenberg E., Nieto O.: Librosa: Audio and music signal analysis in Python. In *Proceedings the Python Science Conference* (Austin, Texas, USA, 2015), pp. 18–25. doi:10.25080/Majora-7b98e3ed-003. 4

[MSS20] Manilow E., Seetharman P., Salamon J.: *Open Source Tools & Data for Music Source Separation*. https://source-separation.github.io/tutorial, 2020. URL: https://source-separation.github.io/tutorial. 2

[Mül21a] MÜLLER M.: An educational guide through the FMP notebooks for teaching and learning fundamentals of music processing. *Signals 2*, 2 (2021), 245–285. doi:10.3390/signals2020018. 4

[Mül21b] MÜLLER M.: *Fundamentals of Music Processing – Using Python and Jupyter Notebooks*, 2nd ed. Springer Verlag, 2021. doi:10.1007/978-3-030-69808-9. 3, 4

[MZ19] MÜLLER M., ZALKOW F.: FMP Notebooks: Educational material for teaching and learning fundamentals of music processing. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)* (Delft, The Netherlands, 2019), pp. 573–580. doi:10.5281/zenodo.3527872. 4

[ÖM22] ÖZER Y., MÜLLER M.: Source separation of piano concertos with test-time adaptation. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)* (Bengaluru, India, 2022), pp. 493–500. 2

[PBD18] PANTELI M., BENETOS E., DIXON S.: A review of manual and computational approaches for the study of world music corpora. *Journal of New Music Research 47*, 2 (2018), 176–189. doi:10.1080/09298215.2017.1418896. 3

[POGS19] PAUWELS J., O'HANLON K., GÓMEZ E., SANDLER M. B.: 20 years of automatic chord recognition from audio. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)* (Delft, The Netherlands, 2019), pp. 54–63. doi:10.5281/zenodo.3527739. 2

[RLS*18] RAFII Z., LIUTKUS A., STÖTER F., MIMILAKIS S. I., FITZGERALD D., PARDO B.: An overview of lead and accompaniment separation in music. *IEEE/ACM Transactions on Audio, Speech, and Language Processing 26*, 8 (2018), 1307–1335. doi:10.1109/TASLP.2018.2825440. 2

[RSM23] ROSENZWEIG S., SCHWERBAUM F., MÜLLER M.: Computer-assisted analysis of field recordings: A case study of Georgian funeral songs. *ACM Journal on Computing and Cultural Heritage (JOCCH) 16*, 1 (2023), 1–16. URL: https://doi.org/10.1145/3551645, doi:doi.org/10.1145/3551645. 3

[SED18] STOLLER D., EWERT S., DIXON S.: Wave-U-net: A multi-scale neural network for end-to-end audio source separation. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)* (Paris, France, 2018), pp. 334–340. 2

[Ser14] SERRA X.: Creating research corpora for the computational study of music: The case of the CompMusic project. In *Proceedings of the AES International Conference on Semantic Audio* (London, UK, 2014). 3

[SG15] SCHLÜTER J., GRILL T.: Exploring data augmentation for improved singing voice detection with neural networks. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)* (Málaga, Spain, 2015), pp. 121–126. 3

[SLI18] STÖTER F., LIUTKUS A., ITO N.: The 2018 Signal Separation Evaluation Campaign. In *Proceedings of the International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA)* (2018), vol. 10891 of *Lecture Notes in Computer Science*, Springer, pp. 293–305. doi:10.1007/978-3-319-93764-9_28. 2

[SM18] SCHREIBER H., MÜLLER M.: A single-step approach to musical tempo estimation using a convolutional neural network. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)* (Paris, France, 2018), pp. 98–105. 2

[SULM19] STÖTER F., UHLICH S., LIUTKUS A., MITSUFUJI Y.: Open-Unmix – A reference implementation for music source separation. *Journal of Open Source Software 4*, 41 (2019). URL: https://doi.org/10.21105/joss.01667, doi:10.21105/joss.01667. 2

[THFK18] THICKSTUN J., HARCHAOUI Z., FOSTER D. P., KAKADE S. M.: Invariances and data augmentation for supervised music transcription. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (Calgary, Canada, 2018), pp. 2241–2245. 3

[Tza09] TZANETAKIS G.: Music analysis, retrieval and synthesis of audio signals MARSYAS. In *Proceedings of the ACM International Conference on Multimedia (ACM-MM)* (Vancouver, British Columbia, Canada, 2009), pp. 931–932. doi:10.1145/1631272.1631459. 4

[Tza14] TZANETAKIS G.: Computational ethnomusicology: A music information retrieval perspective. In *Proceedings of the Joint Conference 40th International Computer Music Conference (ICMC) and 11th Sound and Music Computing Conference (SMC)* (Athens, Greece, 2014), pp. 69–73. 3

[USW*19] UEDA S., SHIBATA K., WADA Y., NISHIKIMI R., NAKAMURA E., YOSHII K.: Bayesian drum transcription based on nonnegative matrix factor decomposition with a deep score prior. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2019), pp. 456–460. 2

[VWK11] VOLK A., WIERING F., KRANENBURG P. V.: Unfolding the potential of computational musicology. In *Proceedings of the International Conference on Informatics and Semiotics in Organisations (ICISO)* (Leeuwarden, The Netherlands, 2011), pp. 137–144. 3

[WCB18] WEHRMANN J., CERRI R., BARROS R. C.: Hierarchical multi-label classification networks. In *Proceedings of the International Conference on Machine Learning (ICML)* (Stockholm, Sweden, 2018), pp. 5225–5234. 3

[WDS*18] WU C.-W., DITTMAR C., SOUTHALL C., VOGL R., WIDMER G., HOCKMAN J., MÜLLER M., LERCH A.: A review of automatic drum transcription. *IEEE/ACM Transactions on Audio, Speech, and Language Processing 26*, 9 (2018), 1457–1483. 2

[WL19] WU Y., LI W.: Automatic audio chord recognition with MIDI-trained deep feature and BLSTM-CRF sequence decoding model. *IEEE/ACM Transactions on Audio, Speech, and Language Processing 27*, 2 (2019), 355–366. doi:10.1109/TASLP.2018.2879399. 2

[WSM20] WEISS C., SCHREIBER H., MÜLLER M.: Local key estimation in music recordings: A case study across songs, versions, and annotators. *IEEE/ACM Transactions on Audio, Speech, and Language Processing 28* (2020), 2919–2932. doi:10.1109/TASLP.2020.3030485. 2, 3

[WZZ*21] WEISS C., ZEITLER J., ZUNNER T., SCHUBERTH F., MÜLLER M.: Learning pitch-class representations from score–audio pairs of classical music. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)* (Online, 2021), pp. 746–753. 2