Twenty-Five Years of MIR Research: Achievements, Practices, Evaluations, and Future Challenges

Geoffroy Peeters, Zafar Rafii, Magdalena Fuentes, Zhiyao Duan, Emmanouil Benetos, Juhan Nam, Yuki Mitsufuji

Twenty-Five Years of MIR Research: Achievements, Practices, Evaluations, and Future Challenges

Geoffroy Peeters[=*], Zafar Rafii[=†], Magdalena Fuentes[=‡], Zhiyao Duan[**],
Emmanouil Benetos[§], Juhan Nam[¶], Yuki Mitsufuji[‖]

[*]LTCI - Télécom Paris, IP-Paris, France [†]Audible Magic, USA [‡]New York University, USA [**]University of Rochester, USA
[§]Queen Mary University of London, UK [¶]KAIST, South Korea [‖]Sony AI, USA

Abstract—**In this paper, we trace the evolution of Music Information Retrieval (MIR) over the past 25 years. While MIR gathers all kinds of research related to music informatics, a large part of it focuses on signal processing techniques for music data, fostering a close relationship with the IEEE Audio and Acoustic Signal Processing Technical Commitee. In this paper, we reflect the main research achievements of MIR along the three EDICS related to music analysis, processing and generation. We then review a set of successful practices that fuel the rapid development of MIR research. One practice is the annual research benchmark, the Music Information Retrieval Evaluation eXchange, where participants compete on a set of research tasks. Another practice is the pursuit of reproducible and open research. The active engagement with industry research and products is another key factor for achieving large societal impacts and motivating younger generations of students to join the field. Last but not the least, the commitment to diversity, equity and inclusion ensures MIR to be a vibrant and open community where various ideas, methodologies, and career pathways collide. We finish by providing future challenges MIR will have to face.**

Index Terms—**Music information retrieval, MIR, review**

I. Introduction

Music Information Retrieval (MIR) is the field that covers all the research topics involved in the understanding, modeling and (more recently) the processing and generation of music[1]. While comparable research existed before (MIR drew from earlier work such as on symbolic music, speech/music discrimination, beat-tracking, or the development of MPEG-7), they were unified under the MIR umbrella in 2000 with the establishment of the first MIR symposium known today as the International Society for Music Information Retrieval (ISMIR) conference and the related ISMIR organisation[2]. Over the years, the contribution of MIR research to the IEEE Audio and Acoustic Signal Processing (AASP) Technical Commitee (TC) has progressively grown through journals and conferences, notably the Transactions on Audio, Speech, and Language Processing (TASLP), the International Conference on Acoustics, Speech, and Signal Processing (ICASSP), and the Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA). While in 2000, MIR was represented within the "Audio and Electroacoustic" TC by only 8 papers in the EDIC [AUD-MUSI] (Applications to Music), it extended in 2006 to two specific EDICS [AUD-ANSY][3] and [AUD-CONT][4], renamed [AUD-MSP][5] and [AUD-MIR][6] in 2010 in the new AASP TC. Since 2013, around 40 MIR papers are presented every year at ICASSP (with a peak of 47 in 2024) which represents a large fraction (over one third) of the total MIR papers published every year. MIR has been in the top biggest groups of papers since 2013 and represented 18.8% of the AASP papers in ICASSP 2024[7]. Starting in 2025, MIR will be represented by three EDICS categories: "Music analysis", "Music signal processing, production, and separation," and "Audio/symbolic-domain music generation," reflecting the field's ongoing development. The development of MIR in the AASP community can also be found through some special issues in IEEE journals, e.g., 2010 Journal of Selected Topics in Signal Processing (JSTSP) [1] and 2019 Signal Processing Magazine (SPM) [2] on music signal processing.

In this paper, we reflect the development of MIR over the last 25 years. We first review its main research achievements along the three EDICS mentioned above in Section II. We then review several successful practices that fuel the rapid growth of MIR, namely the establishment of shared benchmarking frameworks (Section III), the adoption of reproducible and open science practices (Section IV), the active engagement with industry research and products (Section V), and the strong commitment to Diversity, Equity, and Inclusion (DEI) (Section VI). We finish by highlighting future challenges MIR will have to face in Section VII.

II. MIR Research Achievements

Based on input-output relations, MIR research can be roughly categorized into three kinds: *analysis* (mapping signal to labels), *processing* (mapping signal to signal), and *generation* (mapping labels to signal). In this section, we review some significant achievements in each category. Regarding methodology, similar to other research fields, MIR has progressively

[=]Equal contribution

[1] see the MIReS roadmap: https://mires.eecs.qmul.ac.uk/files/MIRES_Roadmap_ver_1.0.0.pdf

[2] https://ismir.net/

[3] [AUD-ANSY] Audio Analysis and Synthesis

[4] [AUD-CONT] Content-Based Audio Processing

[5] [AUD-MSP] Music Signal Analysis, Processing and Synthesis

[6] [AUD-MIR] Music Information Retrieval and Music Language

[7] DCASE 20.8% and Audio and Speech Source Separation 10.0%

shifted from a "knowledge-driven" paradigm (knowledge is provided to the algorithm by the researcher) to a "data-driven" paradigm (knowledge is acquired from the data), from hand-crafted features and models to end-to-end deep neural network approaches, from supervised to self-supervised, from unimodal to multimodal, and from traditional retrieval tasks to generative artificial intelligence (AI) tasks [3]. This paradigm shift has also led to *foundation models* in MIR, i.e., models pre-trained on large amounts of data and subsequently adapted and used for a wide range of downstream tasks, blurring the boundaries of the three categories reviewed in this section. A detailed survey on the current state of the art on music foundation models can be found at [4]. A recent trend, known as *Differentiable Digital Signal Processing*, aims to bridge knowledge-driven and data-driven approaches by integrating differentiable signal processing into deep learning models [5].

A. Music Analysis

The (content-based) music analysis, which extracts features or predicts labels from audio waveforms, has been the major focus at the start of MIR, hence the Retrieval in the name. This was motivated by the vast volume of music available (first mp3 files, then streaming platforms) that posed challenges in retrieving or organizing music [6], [7]. Also, the limitations of the technology at the start (MFCC, GMM, HMM, SVM, ICA, NMF techniques) could not allow the development of convincing unmixing or music generation applications.

Music analysis tasks are often categorized according to the degree of specificity between the query and the retrieved results. *High-specificity* tasks are concerned with identifying the exact song (e.g., Shazam audio fingerprinting) [8] or different versions of the same song (e.g., different performances of the same song or cover songs) [9] or acoustically-similar songs [10]. *Low-specificity* tasks involve annotating songs (auto-tagging) with text-based tags such as genre, mood, and instruments for music retrieval through text queries or for generating playlists [11]. A large set of tasks target the estimation of attributes related to *music notation*: pitch, dominant-melody, multi-pitch, transcription, sequence of chords, key/mode, positions of beats/downbeats, tempo/meter, global structure (verse/chorus positions) or lyrics transcription.

In the 2000s, MIR research was largely inspired by Speech Processing (SP), adopting features like MFCC and the Speech Recognition paradigm (GMM-HMM acoustic model and language model). However, with the rise of deep learning in the 2010s, its focus gradually shifted towards Computer Vision (CV) and Natual Language Processing (NLP) for inspiration. Advances in deep learning have unified the wide range of MIR tasks into the problem of learning an embedding space where similar pairs are close together and dissimilar pairs are far apart taking into account the specificity [12], [13]. More recently, low-specificity tasks have been handled as benchmark downstreams of foundation models [4].

Among the various MIR tasks, pitch and beat estimation were likely the first to attract significant interest and still do today. In the following, we focus on those.

Pitch estimation has been a fundamental topic throughout the entire history of MIR. While pitch estimation for monophonic signals has been viewed as a well-solved problem using traditional signal processing techniques, the presence of noise and reverberation still poses challenges and has led to the development of neural approaches such as CREPE [14]. Self-Supervised Learning (SSL) models, such as SPICE [15] and PESTO [16], have also been introduced to address the lack of annotated training data. Beyond single-pitch detection, a bigger and richer challenge for music signals is Multi-Pitch Estimation (MPE). Multiple survey articles and tutorials have been published on this topic such as [17]. Before the deep learning era, MPE has been a fertile ground for a large variety of ideas and methodologies including traditional signal processing, sparse coding, Non-negative Matrix Factorization (NMF), probabilistic models, and discriminative models. Nowadays, MPE methods are predominantly neural approaches. This, however, has created large discrepancies on model performance between different music styles. Piano transcription has largely been addressed, thanks to several large-scale datasets such as MAPS [18] and MAESTRO [19]. General ensembles, however, have significantly lagged behind.

The analysis of *beat and downbeat tracking*, and rhythm analysis in general, has been another fundamental topic throughout the entire history of MIR [20]. Historically, these tasks were addressed using signal processing techniques or statistical models [21], focusing on handcrafted features such as onset patterns and harmony changes. In recent years, the integration of deep learning techniques has significantly transformed rhythm-related tasks, resulting in substantial improvements in performance [22]. Along with deep learning models, came the reliance on annotated data to train these models, which causes a bottleneck in robustness and generalizability. Recent approaches are addressing this shortcomings by investigating the use of few data [23] and SSL [24].

B. Music Processing

Over the decades, music demixing [25] and mixing [26] have made significant strides, largely due to advances in deep learning. In contemporary music production, demixing is often followed by restoration techniques such as dereverberation [27], bandwidth extension [28], and declipping [29]. Notably, music automixing has reached human-level quality, as evidenced by subjective listening tests [30], thereby opening new possibilities for music remixing and remastering.

C. Music Generation

Music generation has for long been associated with the field of Computer Music (and the related journal and International Computer Music Association (ICMC) conference), mostly focusing on symbolic music. The signal processing community was focusing on sound generation (musical instruments). The development and success of Large Language Models in NLP inspired its translation to the music domain by tokenizing the audio (using VQ-VAE or RVQ) and modeling language using Transformers (such as in OpenAI Jukebox [31], Google MusicLM or Meta MusicGen). Another trend arised from CV with the success of diffusion models for image generation which led to the use of diffusion models to generate spectrograms or the diffusion in the quantized domain -latent-diffusion- (such as in Suno or Stable-audio [32]). Music generation is one of the fastest increasing topics in MIR but also one of the most controversial due to copyright and ethical issues.

III. Benchmarking MIR technologies

The establishment of shared benchmarking frameworks has played an instrumental role in the development of MIR. As early as 2004, the Music Technology Group at UPF proposed the first benchmarking initiative named "Audio Descriptor Contest" [33][8], with the goal of comparing state-of-the-art audio algorithms for a subset of tasks[9]. As for the Detection and Classification of Acoustic Scenes and Events (DCASE), this contest has helped defining what the tasks MIR deals with are, and for each has defined reference training/test sets as well as a set of reference evaluation metrics. This initiative was not sustained but led in 2005 to the Music Information Retrieval Evaluation eXchange (MIREX) [34][10] proposed by the International Music Information Retrieval Systems Evaluation Laboratory (IMIRSEL) who supported until now the effort of benchmarking MIR algorithms.

MIREX has not only allowed to structure MIR around a set of typical tasks but also helped young researchers to quickly put a step in the field. Benchmarking has also motivated the community to put effort on the development of annotated datasets (such as RWC [35] and Isophonics [36]) and to adopt good practices for this (such as for Salami [37]). However, the black-box model used by MIREX[11], along with the absence of a corresponding workshop (such as in DCASE), has limited opportunities for participants to acquire and exchange knowledge, which could partly explain the declining interest in the initiative [38], especially with the development of other initiatives such as the Million Song Dataset Challenge[12] or MediaEval MusiCLEF [39] linked to a workshop. Also, while MIREX was beneficial in MIR's early days, it inherently incited researchers to focus on a fixed set of tasks that progressively became outdated. Only recently were tasks updated to include popular topics like music generation and captioning.

Recently, with the development of large pre-trained models and so-called foundation models, new benchmarking frameworks have been proposed that evaluate each algorithm across numerous MIR tasks. Inspired by SUPERB [40] in speech, evaluation frameworks such as HEAR [41] or MARBLE [42] have been proposed, with the former covering general-purpose audio including music while the latter being music-specific.

IV. Reproducibility and Open Science

Open-source. Another significant part of the field's recent progress can be attributed to the growing adoption of open-source practices. Early on, the creation of the "Matlab Toolbox for MIR" [43] marked a pivotal step, offering a standardized suite of tools. As the demand for more adaptable tools grew, libraries like Essentia [44] emerged, supporting a wide range of MIR applications, from feature extraction to machine learning model integration, and offering browser compatibility. Meanwhile, the Python library `librosa` [45] became a central tool, providing an intuitive and efficient interface for audio and music signal analysis. Other key contributions to the open-source MIR ecosystem include `mir_eval` [46] and `mirdata` [47], which offer standardized evaluation metrics and dataset loaders, streamlining benchmarking and reproducibility in MIR research. Additionally, contributions to MIR from other programming languages such as JAVA [48], and symbolic music tools like `music21` [49], have expanded the scope of MIR tools across different programming environments and applications. The formalization of open-source practices has been promoted through initiatives like "Open-Source Practices for Music Signal Processing Research" [50], which encourage transparency, collaboration, and reproducibility in the broader MIR research community.

Open-access. MIR has also embraced an open-access policy[13] for the publications in the ISMIR conference and the Transactions of MIR (TISMIR) journal.

Open-data. In the early years of MIR, research progress, particularly in data-driven systems, was hindered by limited access to datasets due to copyright restrictions on commercial music. Although this issue remains unresolved, the situation has improved with the creation of purpose-recorded datasets (such as RWC or MedleyDB), the availability of Creative Commons-licensed music (like FMA[14] and MUSDB18[15]), and the workaround of using 7-Digital then YouTube links (as seen with DALI[16] and AudioSet[17]).

Education. Researchers are also highly committed to advancing education in MIR, with tutorials such as [51], foundational texts such as those by Lerch [52] and Müller [53], and the introduction of an "Educational Articles" track in the TISMIR journal [54].

V. MIR Industrial Aspects

The achievements in MIR research have led to successful commercial applications, the creation of startups built around MIR products, and the establishment of strong research and development (R&D) teams working on MIR projects.

One of the first industries which developed with MIR is music identification services. Whether they are consumer apps such as the popular Shazam (acquired by Apple) and

[8] https://ismir2004.ismir.net/ISMIR_Contest.html

[9] In 2004, the tasks were "Genre Classification/Artist Identification," "Melody Extraction," "Tempo Induction," and "Rhythm Classification"

[10] https://www.music-ir.org/mirex/wiki/MIREX_HOME

[11] MIREX uses an "algorithm-to-data" centralized model, where participants submit their algorithms to a central entity that evaluates all systems on private local data, which is not shared afterward.

[12] https://www.kaggle.com/c/msdchallenge

[13] Creative Commons Attribution 4.0

[14] https://github.com/mdeff/fma

[15] https://sigsep.github.io/datasets/musdb.html

[16] https://github.com/gabolsgabs/DALI

[17] https://research.google.com/audioset/index.html

SoundHound, or business-to-business companies such as Audible Magic and Gracenote, their solutions were built upon pioneering audio fingerprinting approaches [8], [55], which have since inspired numerous works, including research on query-by-humming and version identification [9].

The music production industry is another obvious example which benefited from the progress in MIR. Widely-used software, such as Pro Tools[18] from Avid Technologies, Kontakt from Native Instruments, or Ableton Live[19] from Ableton, incorporate a variety of MIR technologies, for example, for beat detection, key analysis, sample categorization, automatic transcription, or noise reduction.

Another notable example of industry which benefited from the progress in MIR is music streaming services. Popular companies such as Pandora (acquired by Sirius XM), Spotify, and Deezer, but also Amazon Music, Apple Music, and YouTube Music from Big Tech, have been extensively relying on MIR solutions, such as music recommendation, music similarity, genre/mood classification, playlist generation, and feature extraction [56]–[58]. In particular, Spotify acquired The Echo Nest in 2014 for $66M[20], a startup created by former MIR academics specialized in the delivery of music content data, which shows the interest for a business around MIR.

More recently, social media platforms, such as YouTube from Google, Instagram from Meta, or TikTok from ByteDance, have also been using MIR technologies, mainly for music identification and recommendation. Their parent companies have well-established R&D teams, such as Google AI[21], Meta's Reality Labs[22], and the Speech, Audio and Music Intelligence (SAMI) team[23] at ByteDance, which work on a variety of projects, including MIR, and regularly publish their works, including at ICASSP, WASPAA, and ISMIR.

The applicability of MIR research can be seen in many other industries, for example, music therapy, music education, music libraries, the gaming industry, the film industry, and more, essentially, wherever music content is being used [52].

VI. Diversity, Inclusion and Societal Impacts

The MIR community has been aware of the lack of underrepresented groups in the field, which is notably evidenced by the gender imbalance observed at ISMIR, although some progress has been seen over the years [59]. In response, the community has been actively working to promote inclusion, with initiatives such as Women in Music Information Retrieval (WiMIR)[24] and its flagship mentoring program. Started in 2011, WiMIR brings together MIR researchers of diverse backgrounds who are dedicated to promoting the role of, and increasing opportunities for women in MIR, through various endeavors. WiMIR has since grown to be inclusive of other minorities and is now striving to promote general DEI in MIR.

ISMIR has also introduced its own mentoring program[25], to encourage newcomers to the conference, by getting more senior members of ISMIR to provide feedback on their papers before submission. Various DEI initiatives are also typically proposed during the conference, such as panel discussions, meetups, and grants to promote the participation of underrepresented communities; the latest ISMIR 2023 has assigned 62 grants for a total budget of more than €40k, which was a prominent effort for a conference with about 350 participants.

To reach out to more people in underrepresented regions such as the Global South, the newly proposed Latin American Music Information Retrieval (LAMIR) workshop[26]will take place in Brazil. Similarly, the AfriMIR[27] initiative was recently launched by the ISMIR board to support, promote, and connect with existing MIR communities across the African continent. ISMIR itself aims to rotate the conference locations between regions of active and emerging MIR research.

The MIR community has also been aware of the cultural inequality in the music being studied; Western classical and pop music have been the dominating genres [60]. Efforts have been made to encourage studies on more diverse genres, such as the special call for papers in ISMIR 2021 and 2022, and a special collection at TISMIR, on cultural diversity in MIR.

VII. Conclusion and future challenges

Over the past 25 years, MIR has emerged as a successful research field, marked by significant technological breakthroughs, a dynamic industrial ecosystem, and a strong community of young researchers who support open science practices and are strongly committed to DEI. The shift from knowledge-driven to data-driven systems, particularly fueled by advancements in deep learning, has not only enhanced performance in established applications, like demixing and source separation, but also enabled the development of new applications such as music generation. However, this brings new challenges for MIR, which are outlined here. While AI has driven substantial progress in many areas, determining how to effectively translate these advancements to MIR remains a key question. Additionally, understanding how these technologies can deepen our knowledge of music is crucial. The significant environmental footprint of training AI systems is a concern. Identifying strategies to mitigate this impact is essential. Although large datasets (like the Million Song Dataset and AudioSet) are available for training data-driven MIR systems, they predominantly reflect Western culture. Preserving cultural diversity within data-driven approaches is a major challenge. Despite major improvements in music demixing and music generation, establishing performance metrics that accurately reflect human perception continues to be problematic. It is anticipated that managing copyrights for generated music will become a significant area of focus in the coming years. Addressing these challenges is essential for the continued growth and positive impact of MIR as a research field.

[18] https://www.avid.com/pro-tools
[19] https://www.ableton.com/en/
[20] https://www.musicbusinessworldwide.com/
[21] https://ai.google/
[22] https://about.meta.com/realitylabs/
[23] https://opensource.bytedance.com/
[24] https://wimir.wordpress.com/
[25] https://ismir2024.ismir.net/new-to-ismir-mentoring-program-2024
[26] https://lamir-workshop.github.io/
[27] https://x.com/afrimir_init?

References

[1] M. Müller e al., “Signal processing for music analysis,” *IEEE J. Sel. Top. Signal Process.*, vol. 5, no. 6, pp. 1088–1110, 2011.

[2] M. Müller et al., “Recent advances in music signal processing,” *IEEE Signal Process. Mag.*, vol. 36, no. 1, pp. 17–19, 2019.

[3] G. Peeters, “The deep learning revolution in MIR: the pros and cons, the needs and the challenges,” in *Proc. of CMMR*, 2019.

[4] M. Yinghao et al., “Foundation models for music: A survey,” *arXiv preprint arXiv:2408.14340*, 2024.

[5] J. H. Engel, L. Hantrakul, C. Gu, and A. Roberts, “DDSP: differentiable digital signal processing,” in *ICLR*, 2020.

[6] M. Casey et al., “Content-based music information retrieval: Current directions and future challenges,” *Proceedings of the IEEE*, vol. 96, no. 4, pp. 668–696, 2008.

[7] M. Schedl, E. Gómez, and J. Urbano, “Music information retrieval: Recent developments and applications,” *Foundations and Trends in Information Retrieval*, vol. 8, no. 2-3, pp. 127–261, 2014.

[8] A. L.-C. Wang, “An industrial strength audio search algorithm,” in *Proc. of ISMIR*, 2003.

[9] F. Yesiler, G. Doras, R. M. Bittner, C. Tralie, and J. Serrà, “Audio-based musical version identification: Elements and challenges,” *IEEE Signal Processing Magazine*, vol. 38, pp. 115–136, 2021.

[10] E. Pampalk, A. Flexer, and G. Widmer, “Improvements of audio-based music similarity and genre classificaton,” in *Proc. of ISMIR*, 2005.

[11] J. Nam, K. Choi, J. Lee, S.-Y. Chou, and Y.-H. Yang, “Deep learning for audio-based music classification and tagging,” *IEEE Signal Processing Magazine*, vol. 36, pp. 41–51, 2018.

[12] S. Chang, D. Lee, J. Park, H. Lim, K. Lee, K. Ko, and Y. Han, “Neural audio fingerprint for high-specific audio retrieval based on contrastive learning,” in *IEEE ICASSP*, 2021.

[13] J. Spijkervet and J. A. Burgoyne, “Contrastive learning of musical representations,” in *Proc. of ISMIR*, 2021.

[14] J. W. Kim, J. Salamon, P. Li, and J. P. Bello, “CREPE: A convolutional representation for pitch estimation,” in *IEEE ICASSP*, 2018.

[15] B. Gfeller, C. Frank, D. Roblek, M. Sharifi, M. Tagliasacchi, and M. Velimirović, “SPICE: Self-supervised pitch estimation,” *IEEE/ACM TASLP*, vol. 28, pp. 1118–1128, 2020.

[16] A. Riou, S. Lattner, G. Hadjeres, and G. Peeters, “PESTO: Pitch estimation with self-supervised transposition-equivariant objective,” in *Proc. of ISMIR*, 2023.

[17] E. Benetos, S. Dixon, Z. Duan, and S. Ewert, “Automatic music transcription: An overview,” *IEEE Signal Processing Magazine*, vol. 36, no. 1, pp. 20–30, 2018.

[18] V. Emiya, R. Badeau, and B. David, “Multipitch estimation of piano sounds using a new probabilistic spectral smoothness principle,” *IEEE TASLP*, vol. 18, no. 6, pp. 1643–1654, 2009.

[19] C. Hawthorne et al., “Enabling factorized piano music modeling and generation with the MAESTRO dataset,” in *ICLR*, 2019.

[20] M. Davies, S. Böck, and M. Fuentes, *Tempo, Beat and Downbeat Estimation*. https://tempobeatdownbeat.github.io/tutorial/intro.html, 2021.

[21] F. Krebs, S. Böck, and G. Widmer, “Rhythmic pattern modeling for beat and downbeat tracking in musical audio.” in *Proc. of ISMIR*, 2013.

[22] S. Böck and M. Davies, “Deconstruct, analyse, reconstruct: How to improve tempo, beat, and downbeat estimation.” in *Proc. of ISMIR*, 2020.

[23] L. S. Maia, M. Rocamora, L. W. Biscainho, and M. Fuentes, “Selective annotation of few data for beat tracking of latin american music using rhythmic features,” *TISMIR*, vol. 7, no. 1, 2024.

[24] D. Desblancs, V. Lostanlen, and R. Hennequin, “Zero-note samba: Self-supervised beat tracking,” *IEEE/ACM TASLP*, vol. 31, pp. 2922–2934, 2023.

[25] Y. Mitsufuji et al., “Music demixing challenge 2021,” *Frontiers in Signal Processing*, vol. 1, 2022.

[26] C. J. Steinmetz, S. S. Vanka, M. A. Martínez Ramírez, and G. Bromham, “Deep learning for automatic mixing,” in *Proc. of ISMIR*, 2022.

[27] K. Saito, N. Murata, T. Uesaka, C. Lai, Y. Takida, T. Fukui, and Y. Mitsufuji, “Unsupervised vocal dereverberation with diffusion-based generative models,” in *IEEE ICASSP*, 2023.

[28] E. Moliner, F. Elvander, and V. Välimäki, “Blind audio bandwidth extension: A diffusion-based zero-shot approach,” *arXiv preprint arXiv:2306.01433*, 2024.

[29] C. Hernandez-Olivan, K. Saito, N. Murata, C. Lai, M. A. M. Ramírez, W. Liao, and Y. Mitsufuji, “VRDMG: vocal restoration via diffusion posterior sampling with multiple guidance,” in *IEEE ICASSP*, 2024.

[30] M. A. M. Ramírez, W. Liao, C. Nagashima, G. Fabbro, S. Uhlich, and Y. Mitsufuji, “Automatic music mixing with deep learning and out-of-domain data,” in *Proc. of ISMIR*, 2022.

[31] P. Dhariwal, H. Jun, C. Payne, J. W. Kim, A. Radford, and I. Sutskever, “Jukebox: A generative model for music,” *CoRR*, vol. abs/2005.00341, 2020. [Online]. Available: https://arxiv.org/abs/2005.00341

[32] Z. Evans et al., “Stable audio open,” *CoRR*, vol. abs/2407.14358, 2024. [Online]. Available: https://doi.org/10.48550/arXiv.2407.14358

[33] P. Cano, E. Gómez, F. Gouyon, P. Herrera, M. Koppenberger, B. Ong, X. Serra, S. Streich, and N. Wack, “ISMIR 2004 audio description contest,” Universitat Pompeu Fabra, Tech. Rep., 2006.

[34] J. S. Downie, “The music information retrieval evaluation exchange (2005-2007): A window into music information retrieval research,” *Acoustical Science and Technology*, vol. 29, no. 4, pp. 247–255, 2008.

[35] M. Goto, “AIST annotation for the RWC music database,” in *Proc. of ISMIR*, 2006.

[36] M. Mauch, C. Cannam, M. Davies, S. Dixon, C. Harte, S. Klozali, D. Tidhar, and M. Sandler, “OMRAS2 metadata project 2009,” in *Proc. of ISMIR*, 2009.

[37] J. B. L. Smith, J. A. Burgoyne, I. Fujinaga, D. De Roure, and J. S. Downie, “Design and creation of a large-scale database of structural annotations,” in *Proc. of ISMIR*, 2011.

[38] G. Peeters, J. Urbano, and G. Jones, “Notes from the ISMIR 2012 late-breaking session on evaluation in music information retrieval,” in *Proc. of ISMIR*, 2012.

[39] C. C. S. Liem, N. Orio, G. Peeters, and M. Schedl, “Brave new task: Musiclef multimodal music tagging,” in *MediaEval*, Pisa, Italy, 10 2012.

[40] S. Yang et al., “SUPERB: Speech processing universal performance benchmark,” in *Interspeech*, 2021.

[41] J. Turian et al., “HEAR: Holistic evaluation of audio representations,” in *NeurIPS*, 2021.

[42] R. Yuan et al., “MARBLE: Music audio representation benchmark for universal evaluation,” in *NeurIPS*, 2023.

[43] O. Lartillot, P. Toiviainen, and T. Eerola, “A matlab toolbox for music information retrieval,” in *Data Analysis, Machine Learning and Applications*. Springer Berlin Heidelberg, 2008.

[44] D. Bogdanov et al., “Essentia: An audio analysis library for music information retrieval,” in *Proc. of ISMIR*.

[45] B. McFee et al., “librosa: Audio and music signal analysis in python,” in *Python in science conference*, 2015, pp. 18–25.

[46] R. Colin et al., “mir_eval: A transparent implementation of common MIR metrics,” in *Proc. of ISMIR*, 2014.

[47] R. M. Bittner, M. Fuentes, D. Rubinstein, A. Jansson, K. Choi, and T. Kell, “mirdata: Software for reproducible usage of datasets,” in *Proc. of ISMIR*, 2019.

[48] C. McKay, J. Cumming, and I. Fujinaga, “JSYMBOLIC 2.2: Extracting features from symbolic music for use in musicological and mir research,” in *Proc. of ISMIR*, 2018.

[49] M. S. Cuthbert and C. Ariza, “music21: A toolkit for computer-aided musicology and symbolic music data,” in *Proc. of ISMIR*, 2010.

[50] B. McFee et al., “Open-source practices for music signal processing research: Recommendations for transparent, sustainable, and reproducible audio research,” *IEEE Signal Processing Magazine*, vol. 36, no. 1, pp. 128–137, 2019.

[51] G. Peeters et al., *Deep Learning 101 for Audio-based MIR, ISMIR 2024 Tutorial*, San Francisco, USA, 2024. [Online]. Available: https://geoffroypeeters.github.io/deeplearning-101-audiomir_book

[52] A. Lerch, *An introduction to audio content analysis: Music Information Retrieval tasks and applications*. John Wiley & Sons, 2022.

[53] M. Müller, *Fundamentals of Music Processing*. Springer Cham, 2015.

[54] M. Müller et al., “Introducing the TISMIR education track: What, why, how?” *Trans. Int. Soc. Music. Inf. Retr.*, vol. 7, no. 1, pp. 85–98, 2024.

[55] J. Haitsma and T. Kalker, “A highly robust audio fingerprinting system,” in *Proc. of ISMIR*, 2002.

[56] J. Joyce, “Pandora and the music genome project, song structure analysis tools facilitate new music discovery,” *Scientific Computing*, vol. 23, no. 14, pp. 40–41, 2006.

[57] K. Jacobson, V. Murali, E. Newett, B. Whitman, and R. Yon, “Music personalization at Spotify,” in *ACM Conference on Recommender Systems*, Boston, MA, USA, September 15-19 2016.

[58] T. Bontempelli et al., “Flow Moods: Recommending music by moods on Deezer,” in *ACM Conference on Recommender Systems*, Seattle, WA, USA, September 18-23 2022.

[59] X. Hu, K. Choi, J. H. Lee, A. Laplante, Y. Hao, S. J. Cunningham, and J. S. Downie, “WiMIR: An informetric study on women authors in ISMIR,” in *Proc. of ISMIR*, 2016.

[60] T. Lidy et al., “On the suitability of state-of-the-art music information retrieval methods for analyzing, categorizing and accessing non-western and ethnic music collections,” *Signal Processing*, vol. 90, no. 4, pp. 1032–1048, 2010.