Renewal Proposal, Deutsche Forschungsgemeinschaft (DFG)

General Information

Name:	**Prof. Dr. rer. nat. Meinard Müller**
Position:	Professor (W3, permanent) für Semantische Audiosignalverarbeitung
Nationality:	German
Institution:	Friedrich-Alexander-Universität Erlangen-Nürnberg
Address (work):	Am Wolfsmantel 33, 91058 Erlangen
Phone:	+49 9131 85-20504
E-mail:	`meinard.mueller@audiolabs-erlangen.de`

Name:	**Dr.-Ing. Jakob Abeßer**
Position:	Senior Scientist
Nationality:	German
Institution:	Fraunhofer Institut für Digitale Medientechnologie (IDMT)
Address (work):	Ehrenbergstraße 31, 98693 Ilmenau
Phone:	+49 3677 467288
E-mail:	`jakob.abesser@idmt.fraunhofer.de`

Topic (initial proposal, MU 2686/11-1, AB 675/2-1): Informed Sound Activity Detection in Music Recordings

Topic (renewal proposal): Informed Sound Activity Detection in Music and Audio Signals

Project Name: ISAD

Subject Area: Computer Science, Artificial Intelligence, Image and Language Processing

Keywords: Music Processing, Audio Processing, Music Information Retrieval

Duration (in months): 36 (Renewal Proposals)

Summary: In music information retrieval (MIR), the development of computational methods for analyzing, segmenting, and classifying music signals is of fundamental importance. In the project's first phase (initial proposal), we explored fundamental techniques for detecting characteristic sound events present in a given music recording. Having a focus on informed approaches that exploit musical knowledge in the form of score information, instrument samples, or representative sections, we considered concrete tasks such as locating audio sections with a specific timbre or instrument, identifying monophonic themes in complex polyphonic music recordings, and classifying music genre or playing styles based on melodic contours. We tested our approaches within complex music scenarios, including Western classical, jazz music, and opera recordings. In the second phase of the project (renewal proposal), our goals will be significantly extended. First, we want to go beyond the music scenario by considering environmental sounds as a second challenging audio domain. As a central methodology, we plan to explore and combine the benefits of model-based and data-driven techniques for learning task-specific sound event representations. Furthermore, we investigate hierarchical approaches that simultaneously incorporate, exploit, learn, and capture sound events that manifest on different temporal scales and belong to hierarchically ordered categories. Considering two distinct audio domains, the general goal of the project's second phase is to gain a deeper understanding of the structural properties of sound events and to obtain explainable deep learning models that are less vulnerable to data biases and confounding factors.

Project Description

1 Starting point

In the following, we report on our main project-related achievements over the last years, which also constitutes state-of-the-art and preliminary work for the renewal proposal. More specifically, we first summarize the objectives and present the main results of the project's first phase (period from November 2017 to July 2021). Then, we give an overview of the project employees and discuss their respective qualifications and contributions. Next, we provide an overview of collaborations, activities, demonstrators, and source code made available in the ISAD project. Subsequently, we discuss deviations from the initial work plan and summarize some follow-up examinations, which also motivate the objectives of the renewal proposal. We refer to the initial proposal for a detailed description of the general state of the art and prior work. We extend this description by discussing the state-of-the-art in acoustic scene classification and environmental sound analysis while summarizing important recent developments in deep learning that form the technical basis for our investigations intended in the second phase of the ISAD project. Finally, in Section 1.2, we list the most important publications of the project's first phase.

1.1 State of the art and preliminary work

1.1.1 Initial objectives and results achieved (initial application)

The general goal of the initial ISAD project was to develop techniques and tools for analyzing, segmenting, and classifying music signals according to various sound characteristics. In particular, we considered detecting specific sound events related to singing voices, certain instruments, or specific musical themes. In the initial proposal, we considered five objectives. In the first three objectives [O1], [O2], and [O3], we considered specific segmentation and classification subtasks related to sound event activity detection. Then we addressed in objective [O4] the general aspect of knowledge integration and covered in objective [O5] various application scenarios. We now summarize these objectives and describe the results achieved in the reporting period from November 2017 to July 2021. Many of the key findings of the ISAD project are based on the publications listed in Section 1.2.

[O1] Detection of Characteristic Sound Events. The presence of a particular instrument or a singing voice is often correlated to specific sound events or characteristic spectro-temporal patterns. In this objective, we studied computational approaches for detecting sound events that are contained in complex sound mixtures.

Achievements:

- [22]
- Structure boundaries in Jazz recordings

[O2] Segmentation of Complex Sound Mixtures. Given a music recording, segmentation may be described as the process of partitioning the audio stream into sections that are of musical/acoustical relevance and that are somehow easier to understand than the original recording. In this objective, our goal was to segment a given music recording into (homogeneous) sections that indicate the presence or absence of certain instruments or other sound events.

Achievements:

- [10]
- [8]
- [20]
- [2]
- [4]

[O3] Sound Event Classification and Automatic Instrument Recognition. Besides *detection* (see [O1]) and *segmentation* (see [O2]), the *classification* of the detected sound events and segments according to musically meaningful categories is another fundamental issue. In this objective, we addressed different classification problems where these categories comprise instruments and singing voices as well as categories on a higher hierarchical level such as instrument families.

Achievements:

- [15]
- [16]
- [3]
- [17]: Harmonic complexity

[O4] Integration of Additional Knowledge. Together with the music recording to be analyzed, one often has additional knowledge on the type of music being played or the sounds to be expected (e.g., note information, number and types of instruments, or sound examples). In this objective, we considered different strategies for incorporating prior knowledge to simplify the segmentation, detection, and classification tasks as specified in the previous objectives.

Achievements:

- [3]
- [16]
- ...

[O5] Applications in Complex Music Scenarios. In this objective, we tested and evaluated our activity detection algorithms by considering various challenging music scenarios including Western classical music, jazz music, and opera recordings.

Achievements:

- [8]: Wagner dataset
- [19]: MTD dataset
- WJD: Jazz dataset with structure annotations; publication in preparation
- [17]

Project employees, qualification of young scientists, contributions

The initial ISAD project was approved for 36 months, providing funds for two full positions—one for Erlangen (FAU) and one for Ilmenau (IDMT). The following list gives an overview of the project's employees as well as the funding periods and volumes (in employee months). Note that most project members were funded only partly by the ISAD-project and partly by other funds (including interruptions due to parental leave, internships, and longer research stays).

- Please carefully revise and extend this list.

- Frank Zalkow (Erlangen, FAU), 01.11.2017 – 31.07.2021 (ca. 50 %, 23 months)
- Michael Krause (Erlangen, FAU), 01.03.2020 – 31.12.2020 (50 %, 5 months)
- Christof Weiß (Erlangen, FAU), 01.09.2018 – 30.06.2019 (50 %, 5 months)
- Stefan Balke (Erlangen, FAU), 01.11.2017 – 30.06.2018 (ca. 50 %, 3 months)
- Michael Tänzer (Ilmenau, IDMT),
- Stylianos Mimilakis (Ilmenau, IDMT),

Besides the project work, the qualification of the next generation of scientists is a central aspect of the ISAD-project. In particular, the PIs see doctoral training as an essential task within academic research. The ISAD-project allowed us to (partly) fund several doctoral students. Additionally, the project also served as a platform for structured doctoral training and joint research that directly connects to the project members' dissertations. The following list indicates the main contributions of the various employees and describes how the project results are linked to their dissertations.

- Please carefully revise and extend this list.
- Frank Zalkow (FAU, 23 months) was the main project employee in Erlangen. He is the main contributor of the research presented in [22, 19, 20]. Mr. Zalkow submitted his PhD Thesis on *Learning Audio Representations for Cross-Version Retrieval of Western Classical Music* in January 2021. In particular, the second part of his dissertation (Part II: Learning Theme-Based Salience Representations, pp. 73–127) is based on the results achieved in the ISAD-project. Mr. Zalkow will defend his dissertation on 28.06.2021.
- Michael Krause (FAU, 5 months) is in the second year of his PhD. He was partly funded by the ISAD-project in his first PhD year, where he made substantial progress in singing voice activity detection, being the main author of [8]. Working on the analysis of opera recordings, Mr. Krause would be an ideal candidate for the second phase of the ISAD-project.
- At the beginning of the project, Stefan Balke (an experienced employee at the end of his doctorate) helped us with preparing two datasets central to the ISAD-project and with training the new project staff and students. His contributions are reflected, among others, by the ISAD-publications [17, 19].
- Michael Tänzer (IDMT) [15] [16]
- Stylianos Mimilakis (IDMT) [10] [9]

Besides supporting doctoral students, the ISAD-project also allowed us to support two young scientists on their way to an academic career. First, Christof Weiß (FAU, 5 months) was funded the ISAD-project in its initial stage, working mainly on music classification tasks (e.g., being the main contributor of [17]). After the opening of his Habilitation procedure in Erlangen (FAU), Dr. Weiß was financed by project-independent state funds (Landesstelle, E14), while continuing collaborating with the doctorate students of the ISAD-project. This has lead to joint publications such as [8, 10, 15]. Second, the ISAD-project allowed Jakob Abeßer (IDMT), the PI in Ilmenau, to further expand his research and teaching profile. Also Dr. Abessser aims for a habilitation at the TU Ilmenau (the procedure being formally opened in ???). In this process, Meinard Müller will serve as a scientific mentor and member of the habilitation committee. At this point, we also want to emphasize that both PIs have not only taken on management and supervision tasks but also actively contributed to the ISAD-project as researchers, which is reflected by several joint publications including [2, 3, 4]. In summary, the ISAD-project has created many synergies and scientific activities beyond the research work conducted by the regular project employees.

Collaborations, activities, demonstrators, and open source code

While advancing state of the art in music and audio processing, one main motivation for having the joint ISAD-project was to strengthen the academic ties between the research groups and the project investigators (PIs) in Erlangen (FAU) and Ilmenau (IDMT). As indicated by the many joint publications and activities, we successfully used the ISAD-project as a platform to foster collaboration on various levels (PIs, PhD students, Master's students) and across different disciplines. Even during the corona crisis, we were able to keep a high level of scientific exchange and personal interaction. In the following, we summarize some of the project's main activities and collaborations triggered by the ISAD-project.

- The following list is ordered by date. Please carefully revise and extend this list.
- **01.–02.12.2017:** Workshop in Illmenau. In this two-day event, PhD students from FAU and IDMT conducted a scientific hackday with topics related to the ISAD-project.
- **23.–24.01.2018:** Project meeting in Erlangen. This two-day event with the PIs and all project members served as a kick-off meeting of the ISAD-project.
- **13.06.2018:** Project meeting in Münchsteinach with the PIs and all project members.
- **28.–29.08.2018:** Project meeting in Ilmenau with the PIs and all project members.
- **07.–09.10.2019:** Research stay by Daniel Stoller (Queen Mary University of London) in Erlangen. Collaboration with Frank Zalkow and Meinard Müller.
- **04.11.2019:** Tutorial on "Fundamentals of Music Processing: An Introduction Using Python and Jupyter Notebooks" by Meinard Müller and Frank Zalkow offered at the International Society for Music Information Retrieval (ISMIR) Conference.[1]
- **09.01.2020:** Project Meeting in Erlangen with the applicants and all project members.
- **21–22.09.2020:** Project meeting (virtual format) with the applicants and all project members.
- **??–??.??.????:** Research stay by Michael Tänzer (IDMT) in Erlangen. Collaboration with Christof Weiss and Meinard Müller.
- **??–??.??.????:** Research stay by Christon-Ragavan Nadar (IDMT) in Erlangen. Collaboration with Christof Weiß, Michael Krause, and Meinard Müller.
- **??–??.??.????:** Research stay by Christof Weiss (FAU) in Ilmenau. Collaboration with Jakob Abeßer and Michael Tänzer.

The PIs attach great importance to the connection between research and teaching. By integrating topics and research results from the ISAD-project, the PI advanced, created, and offered courses such as *Music Processing*[2], *Selected Topics in Deep Learning for Audio, Speech, and Music Processing*[3], and *Machine Listening for Music and Sound Recognition*[4]. These lectures provide essential foundations of music and audio processing and introduce students to current research topics, which are then further deepened in research internships and Master/Bachelor theses. The following list gives examples of student work that is directly related to the ISAD-project and has been supervised by the PIs and the project members.

- The following list is ordered by date. Please carefully revise and extend this list.
- Leo Brütting: Hierarchical Tonal Analysis of Music Signals. Bachelor Thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), March 2019.
- Julian Reck: Boundary Detection in Music Recordings Using Deep Learning Techniques. Master Thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), July 2019.
- David Kopyto: Graph-Based Techniques for Music Structure Analysis of Audio Recordings. Master Thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), March 2020.
- Christon-Ragavan Nadar:
- ...

For many publications, we provided accompanying websites with additional material in the form of freely available audio samples, visualizations, and sonifications. For demonstration purposes,

[1] https://ismir2019.ewi.tudelft.nl/?q=tutorials

[2] https://www.audiolabs-erlangen.de/fau/professor/mueller/teaching

[3] https://www.audiolabs-erlangen.de/fau/professor/mueller/teaching/2021s_dla

[4] https://machinelistening.github.io/

some websites also integrate web-based interfaces, which allow users to access, comprehend, interact, and evaluate the data and results. The implementation of interfaces and maintenance of websites, a labor- and time-intensive work, was conducted by the project members and student assistants (HIWIs). Furthermore, research code of the ISAD-project has been published under suitable open-source licenses. The following list gives an overview of freely available web-based sources and demonstrators of the ISAD project:

- Results [20]: `https://www.audiolabs-erlangen.de/resources/MIR/2019-ICASSP-BarlowMorgenstern`
- MTD Dataset [19]: `https://www.audiolabs-erlangen.de/resources/MIR/MTD`
- Code (annotation tool) [19]: `https://github.com/fzalkow/mtd-alignment-tool`
- Results [22]: `https://www.audiolabs-erlangen.de/resources/MIR/2020-ISMIR-ctc-chroma`
- Teaching material [11]: `https://audiolabs-erlangen.de/FMP`
- Code (teaching material) [11]: `https://github.com/meinardmueller/libfmp`

Deviations from initial work plan and follow-up examinations

The initial ISAD project was approved for 36 months funding two full-time employees. In our initial proposal, we set ourselves five objectives [O1] to [O5], which were approached in 14 work packages. As described before, substantial progress was made in all areas, while the work packages provided a practical guide to the project. In the following, we discuss shifts and deviations from the original work plan and indicate some follow-up examinations, which also motivate the renewal proposal.

Text

State of the art in environmental sound analysis and deep learning (renewal proposal)

In the initial proposal, we provided an overview of the state-of-the-art for various music processing problems that are relevant to the ISAD project, including activity detection, instrument recognition, structure analysis, knowledge integration, and datasets. In the following, we extend this description by discussing the state-of-the-art in acoustic scene classification and environmental sound analysis while summarizing important recent developments in deep learning that form the technical basis for our investigations intended in the second phase of the ISAD project.

State of the art in environmental sound analysis, [1].

As in general multimedia processing, many of the recent advances in MIR have been driven by techniques based on deep learning (DL). For example, DL-based techniques have led to significant improvements for numerous MIR tasks including music source separation [7, 12, 14, 13] music transcription [5], or melody estimation [6]. This trend has been reinforced by the availability of open-source software libraries that allow users (in academia and industry) to develop, implement, and optimize deep neural networks without requiring sophisticated programming and engineering skills. Nowadays, it seems that attaining state-of-the-art solutions via machine learning depends more on the availability of large quantities of data rather than the sophistication of the approach itself [14].

A particular strength of DL-based approaches is their capability of extracting complex features directly from raw audio data, which can then be used for making predictions based on hidden structures and relations []. Furthermore, powerful software packages allow for easily designing, implementing, and experimenting with machine learning algorithms based on deep neural networks (DNNs).

1.2 Project-related publications

In the following, we provide ten published articles with peer review (see Section 1.1.1). These articles reflect the central work done in the first phase of the ISAD project. Further project-related publications, which could not be listed due to the specified maximum number of ten articles, can be found on the website of Meinard Müller[5] and Jakob Abesser[6].

[3] Jakob Abeßer and Meinard Müller. Fundamental frequency contour classification: A comparison between hand-crafted and CNN-based features. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 486–490, Brighton, UK, 2019.

[8] Michael Krause, Meinard Müller, and Christof Weiß. Singing voice detection in opera recordings: A case study on robustness and generalization. *Electronics*, 10(10):1214:1–14, 2021.

[9] Stylianos Ioannis Mimilakis, Konstantinos Drossos, Estefanía Cano, and Gerald Schuller. Examining the mapping functions of denoising autoencoders in singing voice separation. *IEEE/ACM Transactions on Audio, Speech & Language Processing*, 28:266–278, 2019.

[10] Stylianos I. Mimilakis, Christof Weiß, Vlora Arifi-Müller, Jakob Abeßer, and Meinard Müller. Cross-version singing voice detection in opera recordings: Challenges for supervised learning. In *Machine Learning and Knowledge Discovery in Databases – Proceedings of the International Workshops of ECML PKDD 2019, Part II*, volume 1168 of *Communications in Computer and Information Science*, pages 429–436, Würzburg, Germany, 2019.

[15] Michael Taenzer, Jakob Abeßer, Stylianos I. Mimilakis, Christof Weiß, Hanna Lukashevich, and Meinard Müller. Investigating CNN-based instrument family recognition for Western classical music recordings. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 612–619, Delft, The Netherlands, 2019.

[16] Michael Taenzer, Stylianos I. Mimilakis, and Jakob Abeßer. Informing piano multi-pitch estimation with inferred local polyphony based on convolutional neural networks. *Electronics*, 10, 2021.

[17] Christof Weiß, Stefan Balke, Jakob Abeßer, and Meinard Müller. Computational corpus analysis: A case study on jazz solos. In *Proceedings of the 19th International Society for Music Information Retrieval Conference (ISMIR)*, pages 416–423, Paris, France, 2018.

[19] Frank Zalkow, Stefan Balke, Vlora Arifi-Müller, and Meinard Müller. MTD: A multimodal dataset of musical themes for MIR research. *Transactions of the International Society for Music Information Retrieval (TISMIR)*, 3(1):180–192, 2020.

[20] Frank Zalkow, Stefan Balke, and Meinard Müller. Evaluating salience representations for cross-modal retrieval of Western classical music recordings. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 331–335, Brighton, UK, 2019.

[22] Frank Zalkow and Meinard Müller. Using weakly aligned score–audio pairs to train deep chroma models for cross-modal music retrieval. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 184–191, Montréal, Canada, 2020.

2 Objectives and work programme

2.1 Anticipated total duration of the project

36 months

2.2 Objectives

In the first phase of the ISAD project (initial proposal), we had five objectives [O1] to [O5], which are listed here again for convenience. While these objectives form the starting point for our future

[5] https://www.audiolabs-erlangen.de/fau/professor/mueller/publications

[6] https://www.idmt.fraunhofer.de/en/institute/doctorands/abesser.html#1

investigations, we significantly extend our goals for the second phase of the ISAD project with respect to the audio scenarios, the methodology, and applications. The renewal proposal consists of four new objectives referred to as [O6], [O7], [O8], and [O9]. Better motivate why we consider the following objectives. Why are they important in our context? How does this continue our previous research in a natural way? Add one motivating sentence for each new objective. With objective [O6], we go beyond the music scenario by considering environmental sounds as a second challenging audio domain. In [O7], we explore and combine the benefits of model-based and data-driven techniques for learning task-specific sound event representations. Furthermore, in [O8], we will consider hierarchical approaches that simultaneously incorporate, exploit, learn, and capture sound events that manifest on different temporal scales and belong to hierarchically ordered categories. Considering two distinct audio domains, our general and overarching objective [O9] is to gain a deeper understanding of sound events' structural properties and obtain explainable deep learning models, thus contributing to fundamental research of practical relevance.

[O1] **Detection of Characteristic Sound Events.**

[O2] **Segmentation of Complex Sound Mixtures.**

[O3] **Sound Event Classification and Automatic Instrument Recognition.**

[O4] **Integration of Additional Knowledge.**

[O5] **Applications in Complex Music Scenarios.**

[O6] **Generalization Towards other Audio Domains.**

[O7] **Representation Learning.**

[O8] **Hierarchical Approaches.**

[O9] **Towards Sound Event Understanding.**

Similar to its first phase, we expect that the second phase of the ISAD project will support the staff members for pursuing a PhD. Furthermore, it will allow Dr. Abeßer in his role as PI to broaden his research field while advancing his habilitation at the TU Ilmenau. Finally, the ISAD project has the scientific and educational potential for various Master and Bachelor thesis projects in computer science and engineering.

2.3 Work programme including proposed research methods

The following text comes from the initial proposal and needs to be adapted. The following work programme is structured into fourteen work packages, which are to be handled by the two full scientific staff members in a parallel, interlocked, and collaborative fashion. The estimated time required to handle each of the specified work packages is roughly four to six months. Based on each groups' expertise, the work packages (WE**??**) to (WE**??**) are thought for Erlangen and the work packages (WI**??**) to (WI**??**) for Ilmenau. However, this partitioning is not meant to be strict, and a close collaboration throughout the project is highly supported. An intensive cooperation is also ensured by having the integrative work packages (WA**??**) to (WA**??**), where we consider various concrete application scenarios. For an overview of the temporal arrangement of the work packages, we refer to Figure 1 at the end of this section.

[W1] **Text.** Description.

[W2] **Text.** Description.

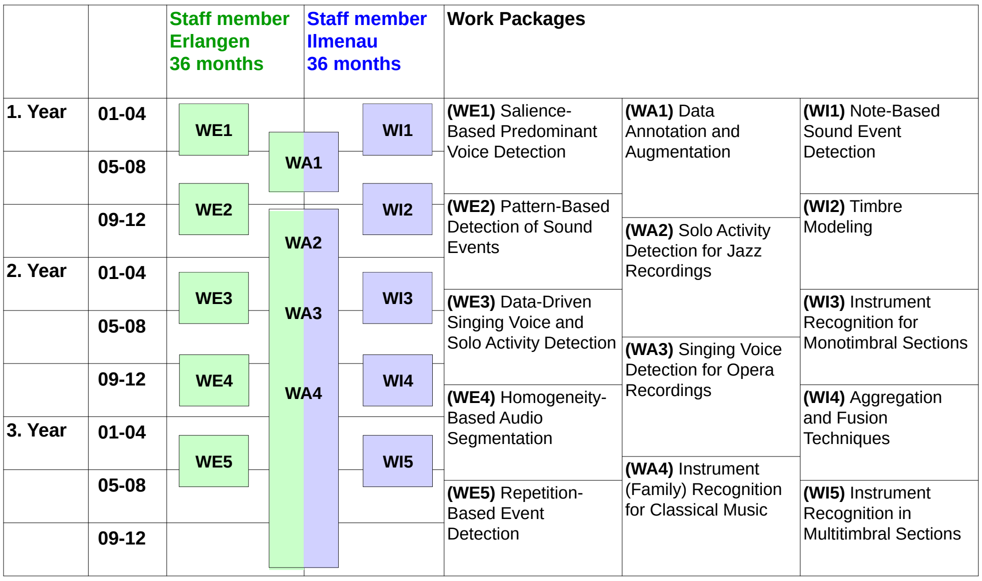
Staff member Erlangen 36 months
Staff member Ilmenau 36 months
Work Packages
1. Year
01-04
05-08
09-12
2. Year
01-04
05-08
09-12
3. Year
01-04
05-08
09-12
WE1
WE2
WE3
WE4
WE5
WA1
WA2
WA3
WA4
WI1
WI2
WI3
WI4
WI5
(WE1) Salience-Based Predominant Voice Detection
(WE2) Pattern-Based Detection of Sound Events
(WE3) Data-Driven Singing Voice and Solo Activity Detection
(WE4) Homogeneity-Based Audio Segmentation
(WE5) Repetition-Based Event Detection
(WA1) Data Annotation and Augmentation
(WA2) Solo Activity Detection for Jazz Recordings
(WA3) Singing Voice Detection for Opera Recordings
(WA4) Instrument (Family) Recognition for Classical Music
(WI1) Note-Based Sound Event Detection
(WI2) Timbre Modeling
(WI3) Instrument Recognition for Monotimbral Sections
(WI4) Aggregation and Fusion Techniques
(WI5) Instrument Recognition in Multitimbral Sections

Figure 1. This is the overview from the initial proposal, which needs to be replaced. Overview of the rough temporal arrangement of the working packages of the second phase of the ISAD project (renewal proposal).

2.4 Data Handling

We need to overwork the following paragraph. We may also mention github and zenodo. In view of reproducibility issues, we mainly want to use data resources that are publicly available. This particularly holds for the new datasets for the research on environmental sounds. As in the first phase of the ISAD-project, we will make additional annotations as well as segmentation/detection results publicly available on a suitable project website. Furthermore, we plan to have publicly accessible websites with demonstrators, numerous sound examples, and suitable source code.

3 Bibliography concerning the state of the art, the research objectives, and the work programme

[1] J. Abeßer, *A review of deep learning based methods for acoustic scene classification*, Applied Sciences, 10 (2020).

[2] J. Abeßer, S. Balke, and M. Müller, *Improving bass saliency estimation using label propagation and transfer learning*, in Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), Paris, France, 2018, pp. 306–312.

[3] J. Abeßer and M. Müller, *Fundamental frequency contour classification: A comparison between hand-crafted and CNN-based features*, in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Brighton, UK, 2019, pp. 486–490.

[4] ———, *Jazz bass transcription using a u-net architecture*, Electronics, 10 (2021).

[5] E. Benetos, S. Dixon, Z. Duan, and S. Ewert, *Automatic music transcription: An overview*, IEEE Signal Processing Magazine, 36 (2019), pp. 20–30.

[6] R. M. Bittner, B. McFee, J. Salamon, P. Li, and J. P. Bello, *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.

[7] I. Kavalerov, S. Wisdom, H. Erdogan, B. Patton, K. W. Wilson, J. L. Roux, and J. R. Hershey, *Universal sound separation*, in Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), New Paltz, New York, USA, 2019, pp. 175–179.

[8] M. Krause, M. Müller, and C. Weiss, *Singing voice detection in opera recordings: A case study on robustness and generalization*, Electronics, 10 (2021), pp. 1214:1–14.

[9] S. I. Mimilakis, K. Drossos, E. Cano, and G. Schuller, *Examining the mapping functions of denoising autoencoders in singing voice separation*, IEEE/ACM Transactions on Audio, Speech & Language Processing, 28 (2019), pp. 266–278.

[10] S. I. Mimilakis, C. Weiss, V. Arifi-Müller, J. Abeßer, and M. Müller, *Cross-version singing voice detection in opera recordings: Challenges for supervised learning*, in Machine Learning and Knowledge Discovery in Databases – Proceedings of the International Workshops of ECML PKDD 2019, Part II, vol. 1168 of Communications in Computer and Information Science, Würzburg, Germany, 2019, pp. 429–436.

[11] M. Müller and F. Zalkow, *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.

[12] Z. Rafii, A. Liutkus, F. Stöter, S. I. Mimilakis, D. FitzGerald, and B. Pardo, *An overview of lead and accompaniment separation in music*, IEEE/ACM Transactions on Audio, Speech, and Language Processing, 26 (2018), pp. 1307–1335.

[13] D. Stoller, S. Ewert, and S. Dixon, *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.

[14] F. Stöter, S. Uhlich, A. Liutkus, and Y. Mitsufuji, *Open-Unmix – A reference implementation for music source separation*, Journal of Open Source Software (JOSS), 4 (2019), p. 1667.

[15] M. Taenzer, J. Abeßer, S. I. Mimilakis, C. Weiss, H. Lukashevich, and M. Müller, *Investigating CNN-based instrument family recognition for Western classical music recordings*, in Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), Delft, The Netherlands, 2019, pp. 612–619.

[16] M. Taenzer, S. I. Mimilakis, and J. Abeßer, *Informing piano multi-pitch estimation with inferred local polyphony based on convolutional neural networks*, Electronics, 10 (2021).

[17] C. Weiss, S. Balke, J. Abeßer, and M. Müller, *Computational corpus analysis: A case study on jazz solos*, in Proceedings of the 19th International Society for Music Information Retrieval Conference (ISMIR), Paris, France, 2018, pp. 416–423.

[18] C. Weiss, F. Brand, and M. Müller, *Mid-level chord transition features for musical style analysis*, in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Brighton, UK, 2019, pp. 341–345.

[19] F. Zalkow, S. Balke, V. Arifi-Müller, and M. Müller, *MTD: A multimodal dataset of musical themes for MIR research*, Transactions of the International Society for Music Information Retrieval (TISMIR), 3 (2020), pp. 180–192.

[20] F. Zalkow, S. Balke, and M. Müller, *Evaluating salience representations for cross-modal retrieval of Western classical music recordings*, in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Brighton, UK, 2019, pp. 331–335.

[21] F. Zalkow and M. Müller, *Vergleich von PCA- und Autoencoder-basierter Dimensionsreduktion von Merkmalssequenzen für die effiziente Musiksuche*, in Proceedings of the Deutsche Jahrestagung für Akustik (DAGA), Munich, Germany, 2018, pp. 1526–1529.

[22] ———, *Using weakly aligned score–audio pairs to train deep chroma models for cross-modal music retrieval*, in Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), Montréal, Canada, 2020, pp. 184–191.

4 Relevance of sex, gender and/or diversity

Both research groups in Erlangen (FAU) and Ilmenau (IDMT) actively strive for diversity concerning various dimensions, including gender, cultural background, and research disciplines. Assuming an open, international, and interdisciplinary perspective is a core element of this project, which is also reflected by applicants' international research and teaching activities and cross-disciplinary collaborations (including computer science, electrical engineering, cognitive sciences, musicology, acoustics).

5 Supplementary information on the research context

5.1 Ethical and/or legal aspects of the project

Not applicable.

5.1.1 Ethical and/or legal aspects of the project

Not applicable.

5.1.2 Descriptions of proposed investigations involving experiments on humans or human materials

Not applicable.

5.1.3 Descriptions of proposed investigations involving experiments on animals

5.1.4 Descriptions of projects involving genetic resources (or associated traditional knowledge) from a foreign country

Not applicable.

5.1.5 Descriptions of investigations involving dual use research of concern, foreign trade regulations

Not applicable.

5.2 Data handling

In view of reproducibility issues, we will mainly use data resources that are publicly available. As in the first phase, additional annotations as well as segmentation/detection results produced on these datasets within the ISAD project will be made publicly available on a suitable project website. Furthermore, we plan to have publicly accessible websites with demonstrators, numerous sound examples, and suitable source code.

5.3 Other information

Not applicable.

6 People/collaborations/funding

6.1 Employment status information

Müller, Meinard, Universitätsprofessor (W3), permanent (unbefristet)
Abeßer, Jakob, Senior Scientist (Fraunhofer IDMT, TVÖD 13) permanent (unbefristet)

6.2 First-time proposal data

Not applicable.

6.3 Composition of the project group

Research Group in Erlangen:

- Meinard Müller, Prof. Dr. rer. nat. (W3, FAU)
- Christof Weiß, Dr.-Ing., Postdoc (Wiss. Angestellter, TVL E14, FAU)
- Sebastian Rosenzweig, PhD student (Wiss. Angestellter, TVL E13, FAU)
- Michael Krause, PhD student (Wiss. Angestellter, TVL E13, FAU)
- Yigitcan Özer, PhD student (Wiss. Angestellter, TVL E13, FAU)

In Erlangen, the following people are expected to contribute to the ISAD project: the project-funded staff member (100%, 36 months, intended for Michael Krause and a new PhD student), Prof. Meinard Müller (10%, supervision, organization, project work), Sebastian Rosenzweig (5%, project work, synergies to his work on fundamental frequency estimation), Dr. Christof Weiß (5%, project work and supervision, synergies to his work on tonal analysis), Yigitcan Özer (5%, project work, synergies to his work on audio decomposition and synthesis). Furthermore, the infrastructure of the International Audio Laboratories Erlangen can be used (Dr. Stefan Turowski, coordinator; Elke Weiland, secretary).

Research Group in Ilmenau:

- Jakob Abeßer, Dr.-Ing., TVÖD 13 (Wiss. Mitarbeiter, Fraunhofer IDMT)
- Hanna Lukashevich, Head of *Semantic Music Technologies* (Fraunhofer IDMT)
- Sascha Grollmisch, PhD student (Wiss. Mitarbeiter, Fraunhofer IDMT)
- David-Scott Johnson, Dr., Post-Doc (Wiss. Mitarbeiter, Fraunhofer IDMT)
- Stylianos Ioannis Mimilakis, PhD student (Wiss. Mitarbeiter, Fraunhofer IDMT)
- Michael Taenzer, PhD student (Wiss. Mitarbeiter, Fraunhofer IDMT)
- Christon Ragavan Nadar, PhD student (Wiss. Mitarbeiter, Fraunhofer IDMT)

In Ilmenau (*Semantic Music Technologies* Group at Fraunhofer IDMT), the following people are expected to contribute to the ISAD project: the project-funded staff member (100%, 36 months, intended for Michael Taenzer and Christon Ragavan Nadar), Dr. Jakob Abeßer (10%, supervision, organization, project work), Hanna Lukashevich (5%, project work, synergies to her work on machine learning for music information retrieval), Sascha Grollmisch and Dr. David-Scott Johnson (5%, synergies to their work on sound event detection and industrial sound analysis), Stylianos Ioannis Mimilakis (5%, project work, synergies to his work on source separation). Furthermore, the infrastructure of the Fraunhofer IDMT Ilmenau can be used.

6.4 Researchers in Germany with whom you have agreed to cooperate on this project

ToDo:

Prof. Gerald Schuller (TU Ilmenau)
Senior Prof. Karlheinz Brandenburg (TU Ilmenau)
Prof. Rainer Kleinertz (Institut für Musikwissenschaft, Universität des Saarlandes)
Prof. Martin Pfleiderer (Hochschule für Musik, Weimar)
Prof. Sebastian Stober (IKS-AiLab, Otto-von-Guericke-Universität Magdeburg)

6.5 Researchers abroad with whom you have agreed to cooperate on this project

ToDo:

Dr. Emmanouil Benetos (Machine Listening Lab, Queen Mary University, London, UK)
Prof. Tuomas Virtanen (Audio Research Group, Uni Tampere, Finland)
Prof. Geoffroy Peeters (Télécom ParisTech, Paris, France)

6.6 Researchers with whom you have collaborated scientifically within the past three years

ToDo:

Dr. Stefan Balke (pmOne Group)
Prof. Juan P. Bello (New York University)
Dr. Estefanía Cano (AudioSourceRe)
Prof. Dr. Simon Dixon (Queen Mary University of London, UK)
Dr. Klaus Frieler (HfM Weimar)
Prof. Rainer Kleinertz (Universität des Saarlandes)
Prof. Alexander Lerch (Georgia Institute of Technology, US)
Prof. Martin Pfleiderer (HfM Weimar)
Prof. Frank Scherbaum (Universität Potsdam)
Prof. Vesa Välimäki (Aalto University, Finland)
Prof. Sebastian Stober (Otto-von-Guericke-Universität Magdeburg)
Prof. Gerhard Widmer (Johannes Kepler University Linz, Austria)

6.7 Project-relevant cooperation with commercial enterprises

Not applicable.

6.8 Project-relevant participation in commercial enterprises

Not applicable.

6.9 Scientific equipment

All scientific equipment required for this project (hardware, software) is available

6.10 Other submissions

No other application for funding of this project has been submitted. If we make such a proposal, we will immediately inform the German Research Foundation.

7 Requested modules/funds

7.1 Basic module

7.1.1 Funding for staff

Scientific staff (full position, TVL 13, 36 months, Erlangen). The staff member takes over the work for Erlangen throughout the total duration of the project. She/he needs to have excellent qualifications in the areas of Music Information Retrieval, Digital Signal Processing, and Machine Learning. Furthermore, comprehensive programming skills in Python are required. Basic musicological knowledge is requested. A highly qualified candidate for this position is Michael Krause, who already worked in the first phase of the ISAD-project. The DFG position would allow him to concentrate on his research while finishing his PhD (third and fourth year). Furthermore, we plan to hire a new PhD student joining the ISAD-project at the beginning of her/his PhD.

Scientific staff (full position, TVÖD 13, 36 months, Ilmenau). The staff member takes over the work in Ilmenau throughout the total duration of the project. She/he needs to have excellent qualifications in the areas of Audio Signal Processing and Machine Learning with a focus on Deep Learning. Furthermore, very good programming skills in Python including prior knowledge of common machine learning libraries such as scikit-learn, keras, tensorflow, or pytorch are required.

Four student assistants (à 36 months à 30h/month, Erlangen and Ilmenau). To support the staff members, we ask for four student assistants (two for Erlangen and two for Ilmenau) with excellent programming skills (in particular Python and/or Java) and a good musical background. The student assistants are required for several research-related tasks. First, prototypical interfaces and web-based demonstrators for the applications are to be developed and implemented. Second, standard signal processing and machine learning algorithms from the scientific literature need to be (re-)implemented, adapted, and tested. Third, datasets and annotations need to be maintained and generated. Fourth, evaluations and tests are to be conducted. One superordinate goal of the ISAD project is to introduce motivated students to current research problems at an early stage of their studies. This can be achieved by integrating them into the research group by means of student assistant positions, which may then lead to Master and Bachelor thesis projects related to the ISAD project.

7.1.2 Direct Project Costs

7.1.2.1 Equipment up to EUR 10,000, Software and Consumables. Not applicable.

7.1.2.2 Travel Expenses. The expected research results should be published and presented at major international conferences (e.g., ICASSP, ISMIR, EUSIPCO, ACM Multimedia, WASPAA, AES). Per year, we expect that each of the two staff members visits two major conferences (with an estimated cost of EUR 2000 per conference comprising overseas flight, hotel, and registration fee) and several smaller conferences, workshops, and visits of the project partners. This amounts to an estimated cost of EUR 5000 per year and per staff member. Over three years, this results in the following estimation for travel expenses: EUR 30000

7.1.2.3 Visiting Researchers (excluding Mercator Fellows). We would like to invite at least one guest scientist per year per project partner for a guest talk and joint research (estimated at 600 EUR per visit), resulting in a total amount of: EUR 3600

7.1.2.4 Expenses for Laboratory Animals. Not applicable.

7.1.2.5 Other Costs. Not applicable.

7.1.2.6 Project-Related Publication Expenses. Open-access articles in renowned peer-reviewed journals are subject to a fee (between 500 and 1500 EUR per article). To this end, we request the following allowance: EUR 6000

7.1.3 Instrumentation

To be discussed.

7.2 Module workshop funding

To be discussed.

7.3 Module public relations funding

To be discussed.