A decade of DCASE: Achievements, practices, evaluations and future challenges

Annamaria Mesaros[1], Romain Serizel[2], Toni Heittola[1], Tuomas Virtanen[1], Mark D. Plumbley[3]
[1] *Signal Processing Research Center, Tampere University*, Tampere, Finland
[2] *Université de Lorraine, CNRS, Inria, Loria*, Nancy, France
[3] *Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, UK*
annamaria.mesaros@tuni.fi, romain.serizel@loria.fr , toni.heittola@tuni.fi, m.plumbley@surrey.ac.uk , tuomas.virtanen@tuni.fi

***Abstract*—This paper introduces briefly the history and growth of the Detection and Classification of Acoustic Scenes and Events (DCASE) challenge, workshop, research area and research community. Created in 2013 as a data evaluation challenge, DCASE has become a major research topic in the Audio and Acoustic Signal Processing area. Its success comes from a combination of factors: the challenge offers a large variety of tasks that are renewed each year; and the workshop offers a channel for dissemination of related work, engaging a young and dynamic community. At the same time, DCASE faces its own challenges, growing and expanding to different areas. One of the core principles of DCASE is open science and reproducibility: publicly available datasets, baseline systems, technical reports and workshop publications. While the DCASE challenge and workshop are independent of IEEE SPS, the challenge receives annual endorsement from the AASP TC, and the DCASE community contributes significantly to the ICASSP flagship conference and the success of SPS in many of its activities.**

***Index Terms*—DCASE Challenge, DCASE Workshop, AASP Challenges**

I. Introduction

Early research in the field of acoustic scene analysis and event detection comprises work on computational auditory scene analysis (CASA) which aimed to mimic human auditory perception to segregate and identify sound sources in complex audio environments [1], [2]. Publications focused on various aspects of audio signal processing, such as feature extraction, and pattern recognition algorithms for sound classification, often comparing effectiveness of known features from other domains, like mel-frequency cepstral coefficients (MFCCs) and temporal features like zero crossing rate, spectral centroid, short-time energy for classifying environmental sounds [3] or scenes [4]. Classification techniques were based mainly on support vector machines (SVMs) [3], [4], Gaussian mixture models and hidden Markov models [4].

Prior to the DCASE challenges, there were limited publicly available datasets for training and evaluating acoustic scene or sound recognition systems, most notably RWCP [5]. In terms of challenges, the CLEAR Evaluations on the Classification of Events, Activities, and Relationships conducted in 2006 and 2007 had a task focused on Acoustic Scene Analysis to detect 12 categories of *non-speech noises*; the task had 2 entries in 2006 [6], and 5 in 2007 [7], but there was no further activity after the second edition. Subsequently, studies on environmental sound recognition used in-house datasets and were therefore non-reproducible. Moreover, the choice of evaluation procedures was also up to the author, with no clear agreement on metrics or standardized benchmarks. This lack of standardization hindered the comparability of results across different studies.

The formalization of the DCASE challenges in 2013 [8] marked a significant change. With standardized datasets and evaluation protocols, the DCASE challenges provided a structured platform for researchers to benchmark their algorithms against a common dataset, leading to more reproducible and comparable results. The result was a more collaborative and competitive research environment that fostered significant advancements in the field for acoustic scene classification, sound event detection, and other related tasks.

II. Open science through data challenges

Looking back for a brief history of data evaluation challenges, MIREX [9] or CHiME [10] stand out. MIREX began in 2005 as an annual evaluation campaign for Music Information Retrieval (MIR), evaluating algorithms for tasks like music genre classification, melody extraction, and music similarity. Over time, MIREX expanded to more complex tasks, becoming a benchmark for MIR research. MIREX promoted open-source practices and detailed reporting, emphasizing rigorous evaluation and reproducibility, even though it was not explicitly focused on open science as we know it today. The CHiME challenges began in 2011, focusing on robust automatic speech recognition (ASR) in noisy environments, with typical tasks being speech enhancement, speaker diarization and speech recognition in real-world noisy environments. Through multiple iterations, CHiME has introduced more complex scenarios like distant microphone conversational speech recognition in everyday home environments [10]. DCASE started in 2013 under endorsement of the IEEE Audio and Acoustic Signal Processing (AASP) Technical Committee through its AASP Challenges Subcommittee[1], and became a very successful and far-reaching endeavor.

A. The first DCASE Challenge

Before the first DCASE challenge, a one-day "Machine Listening Workshop" organized at Queen Mary University of London, UK in December 2010 provided an early indication of the interest in research in this area. Subsequent discussions in 2011 between Queen Mary University of London, UK (Benetos, Giannoulis, Stowell, Plumbley) and IRCAM, Paris, France (Rossignol, Lagrange) led to the idea for the first challenge. At around that time, Plumbley had also joined the Audio and Acoustics Signal Processing Technical Committee (AASP TC) of the IEEE Signal Processing Society, which was organizing a new "IEEE AASP Challenge" series. A proposal for a challenge in "Detection and Classification of Acoustic Scenes and Events" was submitted and approved in early 2012, becoming the second "IEEE AASP Challenge". The Chairs of the following year's WASPAA 2013

This work was supported in part by Academy of Finland grant 332063 "Teaching machines to listen", and by the UK Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/T019751/1 "AI for Sound". For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) license to any Author Accepted Manuscript version arising. This publication is supported by datasets that are openly available at locations referenced in this paper.

[1]https://signalprocessingsociety.org/community-involvement/audio-and-acoustic-signal-processing/aasp-challenges

workshop also agreed to support the challenge, through a special poster session with an overview oral presentation.

The first DCASE challenge included two tasks: Acoustic scene classification (ASC), and Sound Event Detection (SED). These tasks have been present in each edition to date, but their setups have evolved to be more complex and closer to real-life applicability, and datasets became significantly larger and more diverse. For the Acoustic Scene Classification task, after considering whether to use existing data, it was decided to collect new data, to avoid inconsistency of microphones and recording equipment. Recordings were made from ten different types of sound scenes in the London area, and were released in 30-second segments, with 10 examples of each of the 10 scenes, totalling 50 minutes of audio. The Sound Event Detection task included two sub-tasks, based on sounds recorded in offices: an "Office Live" (OL) task and an "Office Synthetic" (OS) sub-task. The OL recordings were live recordings of scripted event sequences, with no overlapping events. The OS recordings were synthesized from live recordings of individual events, combined into synthetic mixtures which included ambient background, which allowed different polyphony (overlap) of sound events and different signal-to-noise (SNR) levels of event sounds over background.

B. DCASE Challenges 2016-2024

The DCASE Challenge was restarted in 2016 with support from an ERC Starting Grant by Virtanen, which facilitated data collection and annotation at a larger scale. Initially organized by a core group of people through existing collaborations, it soon became evident that this approach is unsustainable due to its popularity, reflected in large number of participants. Thus, an open call for tasks was introduced.

The first open call for task proposals, published in December 2018, seeked to include new tasks, unrelated to those from the previous years. For 2019, the Steering Group selected five tasks; these included the longstanding ASC and SED, but also new tasks like audio tagging (AT) with noisy labels [11], sound event localization and detection (SELD) [12]. The decentralized organization makes the challenge a community effort, distributing the organization effort to multiple researchers who have full control over their task. The challenge operates under unified coordination, with the challenge coordinators (Mesaros and Serizel) aligning the format of the tasks.

The open call for task organization had a profound impact on the challenge and the entire research field, diversifying the topics of the challenge. New research directions that emerged are sound event localization and detection (SELD), the use of weakly-labeled data [13], and few-shot learning for bioacoustic event detection [14]. Furthermore, the introduction of tasks like automated audio captioning [15] and language-based audio retrieval [16] highlights the growing intersection of audio processing with natural language processing. The latest addition to the DCASE challenge is an audio generation task [17], marking a significant step forward from what was essentially audio signal processing and analysis. As a result, the DCASE research field has seen notable advancements, the diversity in tasks, shown in numbers in Table I, ensuring that it remains vibrant, highly-engaging, and continually evolving.

C. The DCASE Challenge process

The typical timeline of the challenge starts with the open call for task proposals. The review and selection of tasks is done by the Steering Group, with the selected tasks being announced in January. Task organizers are responsible with producing and publishing the open datasets and baseline systems for their task, and evaluating the submissions. The challenge opens in March/April, when complete information on tasks is provided to potential participants: a detailed task description including the evaluation procedure, the development dataset, and the baseline system (code and results according to the task requirements).

TABLE I
DCASE CHALLENGE STATISTICS.

Edition	Research topics	Tasks and Subtasks	Teams	Entries
2023	7	9	123	428
2022	6	7	135	410
2021	6	8	127	393
2020	6	7	138	473
2019	5	7	109	311
2018	5	7	81	223
2017	4	4	74	200
2016	4	4	66	85
2013	2	2	21	31

Participants have about two months to develop their algorithms. The evaluation datasets (when required by the task) are provided 2-4 weeks before the submission deadline. The participants are expected to submit the system outputs for the evaluation data, a technical report describing their method, and additional meta information related to the submission, using provided templates. The notable difference from the first DCASE is the submission of system outputs instead of code: running the code submitted by participants was a very time-consuming process [18], hence the choice of simplifying the evaluation process.

Because the key dates are aligned for all tasks, DCASE challenge feels like one single event, even though there are multiple teams working under its umbrella. Publication of task descriptions, development datasets, submission deadline and publication of results happens for all tasks at once (with small exceptions). The technical reports are published unaltered on the challenge website at the same time with the results.

III. DCASE WORKSHOPS

Dissemination of the challenge results was initially organized during a special session at WASPAA 2013, later as a satellite workshop to EUSIPCO 2016, where many participants attended only the DCASE workshop, not necessarily the main conference. This motivated the regular organisation of a standalone event. Since then, the workshop has been held every year as a companion event to the challenge. Although the workshop and challenge teams operate independently, they maintain communication through the challenge coordinators to ensure their deadlines are compatible.

The DCASE workshop is now a two-and-a-half-day event that attracts around 100 participants each year. Approximately 50 scientific papers are published at each edition, with an acceptance rate of about 50%. The workshop accounts for roughly 15% of the annual publications on DCASE-related topics (included in Fig. 2), making it one of the leading venues for research in this area, alongside ICASSP. In line with the challenge's principle of open science, the workshop proceedings are published under a Creative Commons license.

Due to the close relationship between the workshop and the challenge, the early editions of the workshop primarily featured papers related to challenge submissions, comprising about 80% of the content until 2018. This trend has changed since 2019 and challenge-related papers now make up about half of the workshop's publications. The number of challenge participants increased rapidly from 2013 to 2019 and has since stabilized, as reflected in Table I. A similar trend is observed in workshop attendance (Fig. 1), with the

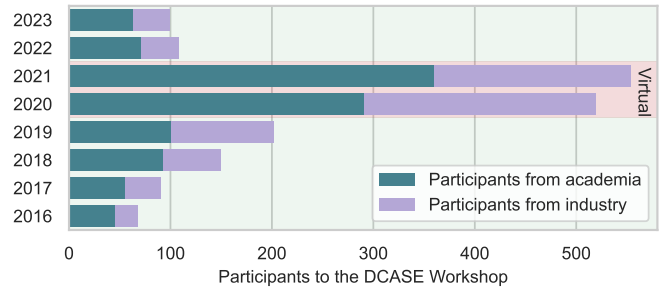

Fig. 1. DCASE Challenge and Workshop participation rates (DCASE 2020 Workshop and DCASE 2021 Workshop were virtual and had free registration

exception of 2020 and 2021, when the workshop was held virtually due to the pandemic and waived registration fees for all participants. Interestingly, the DCASE workshop has attracted significant interest from industry since its inception, with 32% of participants from industry in the first edition. This participation rate has remained stable, averaging 35-38%, and even reaching 50% in 2019. This balance fosters insightful discussions on the relationship between academic research and real-life applications.

IV. DCASE as a research area

The DCASE research area began to flourish once public datasets and clearly defined research questions were introduced through the challenge tasks. The availability of datasets and baseline system code for each task provided researchers with a solid starting point in their work. Figure 2 shows the increase in scientific publications over the years that contain key terms from the DCASE challenge tasks in their title. The tasks within DCASE guide several key research directions, of which a few are briefly introduced below.

Acoustic scene classification aims to classify a test recording into one of the provided predefined classes that characterizes the environment in which it was recorded. Such acoustic scenes include ”airport”, ”public square”, and ”metro”. The ASC task has been active for all ten editions of the DCASE challenge, with changes in the classification setup and the dataset. From a dataset of 50 minutes and 10 classes in 2013 [18], 13 hours and 15 classes in 2017 [19], the largest dataset to date is TAU Urban Acoustic Scenes 2019 [20], consisting of 40 hours of audio and 10 classes, simultaneously recorded with four devices. Research questions tackled multi-device data [21], low-complexity [22], and data-efficient ASC [23].

Sound event detection aims to identify sound event classes and their time boundaries in an audio recording. This task was also present in all the ten editions, with iterative changes. While the task setup was straightforward in the first editions [8], [19], the 2017 edition of the challenge introduced the use of weakly-labeled training data to overcome the strongly labeled data shortage [24]. Since then, the task has evolved to include synthetic training data [13], sound separation as a pre-processing [25], systematic reporting of the computational footprint [26] and the use of soft labels [27] in a framework with potentially missing labels during training [28].

Sound event localization and detection was introduced in 2019, aiming to jointly recognize and localize sound events in 3D space [29]; the task involves detecting the temporal activities of sound events and estimating their spatial locations, using both first-order ambisonics and microphone array recordings. The latest development of the SELD task include using audio-visual input [30] and, in 2024, distance estimation [31].

Bioacoustic sound event detection, a task introduced in 2022, focuses on few-shot learning, which requires algorithms to detect and

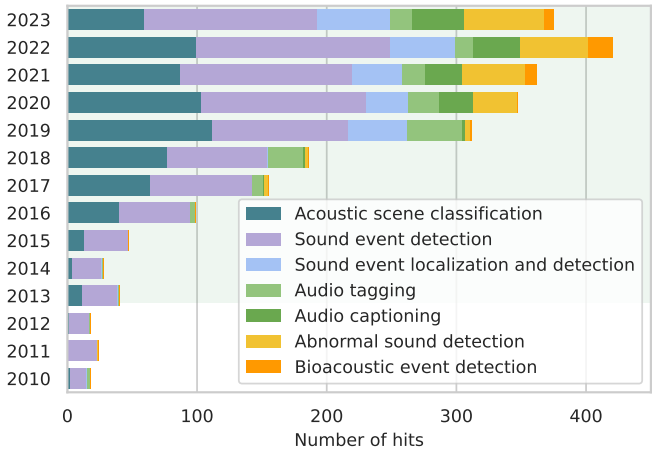

Fig. 2. Number of publications related to DCASE Challenge tasks: Google Scholar hits for papers containing the respective keywords in their title

classify sounds with only a few examples. This approach has evolved over the years, addressing the challenges of data sparsity and class imbalance in bioacoustic datasets [14]. The task's objective is to develop reliable algorithms capable of identifying animal vocalizations in noisy environments with minimal labeled data [32].

Anomalous sound detection focuses on identifying whether a sound emitted from a machine is normal or anomalous, given that only normal sounds are provided for training [33]. The task was first introduced in 2020 and has gained popularity quickly due to its clear industrial applicability. The task has evolved to address practical challenges such as domain shifts [34] and the need for rapid deployment in new environments [35].

Automated audio captioning and **Language-based audio retrieval** are relatively new tasks. Automated audio captioning [36] was introduced in 2021 as the first textual language based task in the DCASE challenge, supplemented the following year with the language-based audio retrieval task [16]. The goal of these tasks is to link the audio and text modalities. Automated audio captioning aims to describe audio recordings in a natural language, beyond sound scene or event classes. Conversely, language-based audio retrieval focuses on finding audio recordings that correspond to a given natural language description from a pool of recordings.

As exemplified above, DCASE has become a diverse research area, expanding from the initial audio analysis to encompass different modalities and tackle real-world applicability scenarios. In addition to the dedicated DCASE challenge and workshop, DCASE is strongly represented in the IEEE SPS conferences and journals; for example there were 65 DCASE-related papers in ICASSP 2024, a number similar to the size of a DCASE workshop.

V. DCASE Community practices and own challenges

DCASE has been successful in building a thriving research community into which a significant number of volunteers are actively contributing. Several factors have enabled building the community:

1) Many researchers in the community shared backgrounds in related audio processing fields like music or speech processing, information retrieval, sound source separation, machine acoustics, and bioacoustics. As a result, they were already well-connected and had enough contacts to gather a critical mass for the new community.

2) Although there was no dedicated funding for building the DCASE community, many individual members and institutions had research grants that enabled them to contribute. These contributions

included organizing and participating in tasks and workshops, managing the web pages, and writing scientific papers and technical reports.

3) DCASE uses a set of communication tools to efficiently disseminate results and discuss within the community. For example, the DCASE web pages (maintained by Heittola from Tampere University) present the tasks in a clear and coherent fashion, with all editions having a consistent look and level of detail. The comprehensive presentation of the results allows for system comparison and analysis beyond the mere ranking. Additionally, the web pages include a significant amount of information about datasets and other resources that have been useful in community building.

4) The community was initially centered around the yearly challenge, which remains an important aspect. The selection of tasks for the challenge over the years has greatly influenced the research directions within the community. The goal in task selection has been to strike a balance between maintaining long-running tasks that attract returning participants and introducing new tasks that open up broader research areas. This outreach to other communities has fostered richer discussions within the community. Task selection and organization have also been crucial for community accessibility. DCASE has ensured that people can participate in the challenge by including entry-level tasks for new participants, including undergraduate students.

5)The DCASE workshop quickly became an established annual event: By 2017, the DCASE Workshop was organized as an independent event, attracting a significant number of participants. The initial workshops in 2016 and 2017 left a positive impression on attendees due to their well-executed organization, including the venue, catering, keynote speakers, and technical program, which helped establish the workshop's credibility as an independent event.

6) Overall, DCASE has maintained high standards for its tasks and workshops, enhancing its credibility within the scientific community, and attracted participation from neighboring fields. Challenge tasks are required to use open data (as explained in Section II). Participants are required to submit technical reports with sufficient documentation of their methods; this way, the challenge contributes to openly sharing knowledge within the community and is not just a competition. The challenge promotes originality and open-source submissions by offering awards determined by a jury composed of task organizers.

7) The community strives for diversity and inclusion. Initially, tasks and workshops were organized by representatives from a limited number of institutes. However, in recent years, this has changed significantly. Now, tasks are organized by geographically diverse teams, including representatives from both academia and industry. DCASE workshop has also experimented with offering grants to encourage the participation of underrepresented minorities or low carbon footprint transportation to the workshop.

8) The DCASE Steering Group has a central role in the community. One of the group's responsibilities is to select the challenge tasks and workshop organization teams, and for this, open calls for task and workshop organisation are published to engage the members of the community. The Steering Group includes representatives from both academia and industry. It is partially renewed every three years, with attention to maintaining a balance of representation from academia and industry, as well as ensuring gender and geographical equity.

A. Challenges and limitations

While DCASE community has been developing successfully regarding many aspects, it has its own challenges to overcome.

The challenges rely heavily on datasets. With the choice of openly available datasets, keeping the tasks interesting and evolving requires regularly collecting and curating data. While it can be fairly easy to find new (small) datasets, ensuring that these datasets allow for some continuity with previous tasks and datasets in order to achieve some mid to long-term support is far from obvious. Balancing between new and old tasks and opening up to new domains when selecting tasks is influenced by the availability of data. This choice can significantly impact the community's work, as it involves a trade-off between topics that easily attract participants and those that, while potentially less attractive, could open new avenues in the domain.

The challenge, like any competition, tends to focus attention on the winner of each task. This brings two main drawbacks. First, a significant portion of the work during the challenge or related to the task is aimed at achieving the highest ranking score. This focus can sometimes hinder originality, as innovative solutions may not be as efficient as well-known, finely-tuned approaches. Second, to ensure fair comparisons between participants, task organizers must design some mechanisms to ensure (not always successfully) that everybody follows the rules. These mechanisms can sometimes limit the scientific aspects explored within the tasks.

Organizing an annual workshop is quite demanding, requiring a team of at least five to ten people for each event. One of the community's challenges is finding new, motivated teams to organize each edition. This is evident in the list of past organizers, which relies on a rather limited pool of individuals who are often already active in the Challenge and the Steering Group.

Communication is a critical aspect in a scientific community. While discussions are very active and fruitful during workshops, they only engage a fraction of the community, as seen when comparing attendance at online and in-person editions. This disparity can be attributed to the various constraints, both time-related and financial, associated with attending in-person workshops. Communication channels were created to overcome this issue, including the DCASE discussion mailing list (since the community's early stages) and a dedicated Slack channel (since the pandemic). However, these tools are primarily active during the challenge period for announcements by organizers or participant requests. Outside of this period, they remain relatively quiet, falling short of achieving their original goal of fostering ongoing discussions within the community.

VI. Future

Within ten years, DCASE has established itself as an important research area within the audio signal processing domain. This expansion was greatly helped by the recurring DCASE challenge and the efforts to steer the community towards open science. Yet, this rapid progression relies on a rather limited pool of volunteers, making the community in need of more active participants.

As the community stabilizes after years of growth, one key question is how to keep it lively and attractive? One approach was to open to other communities to foster broader discussions but this must be done carefully to avoid the risk of dispersion in terms of research themes. Managing this remains an open challenge.

Finally, in a world that is rapidly changing, in particular in a context of climate change, but also with the spreading use of machine learning based systems (which are at the core of our research), it becomes essential to question our environmental and ethical impact. DCASE has addressed this recently by including bio-acoustic tasks, monitoring energy footprints, and setting low computational constraints. Yet this is only the beginning of the changes that need to be made to ensure that the work done in our community is beneficial. This is a key aspect we will have to reflect on in the near future.

REFERENCES

[1] D. P. Ellis, "Prediction-driven computational auditory scene analysis," Ph.D. dissertation, Massachusetts Institute of Technology, 1996.

[2] D. F. Rosenthal, H. G. Okuno, H. Okuno, and D. Rosenthal, *Computational Auditory Scene Analysis: Proc. of the IJCAI-95 Workshop*. CRC press, 2021.

[3] S. Chu, S. Narayanan, and C.-C. J. Kuo, "Environmental sound recognition with time–frequency audio features," *IEEE Trans. Audio Speech Lang. Process.*, vol. 17, no. 6, pp. 1142–1158, 2009.

[4] A. Eronen, V. Peltonen, J. Tuomi, A. Klapuri, S. Fagerlund, T. Sorsa, G. Lorho, and J. Huopaniemi, "Audio-based context recognition," *IEEE Trans. Audio Speech Lang. Process.*, vol. 14, no. 1, pp. 321–329, 2006.

[5] S. Nakamura, K. Hiyane, F. Asano, T. Nishiura, and T. Yamada, "Acoustical sound database in real environments for sound scene understanding and hands-free speech recognition." in *LREC*, 2000.

[6] R. Stiefelhagen, K. Bernardin, R. Bowers, J. Garofolo, D. Mostefa, and P. Soundararajan, "The CLEAR 2006 evaluation," in *Multimodal Technologies for Perception of Humans*, R. Stiefelhagen and J. Garofolo, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 1–44.

[7] R. Stiefelhagen, K. Bernardin, R. Bowers, R. T. Rose, M. Michel, and J. Garofolo, "The CLEAR 2007 evaluation," in *Multimodal Technologies for Perception of Humans*, R. Stiefelhagen, R. Bowers, and J. Fiscus, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 3–34.

[8] D. Stowell, D. Giannoulis, E. Benetos, M. Lagrange, and M. D. Plumbley, "Detection and classification of acoustic scenes and events," *IEEE Trans. Multimedia*, vol. 17, no. 10, pp. 1733–1746, Oct 2015.

[9] J. S. Downie, A. F. Ehmann, M. Bay, and M. C. Jones, *The Music Information Retrieval Evaluation eXchange: Some Observations and Insights*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 93–115.

[10] J. Barker, S. Watanabe, E. Vincent, and J. Trmal, "The fifth CHiME speech separation and recognition challenge: Dataset, task and baselines," 2018. [Online]. Available: https://arxiv.org/abs/1803.10609

[11] E. Fonseca, M. Plakal, D. P. W. Ellis, F. Font, X. Favory, and X. Serra, "Learning sound event classifiers from web audio with noisy labels," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 21–25.

[12] A. Politis, A. Mesaros, S. Adavanne, T. Heittola, and T. Virtanen, "Overview and evaluation of sound event localization and detection in DCASE 2019," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 29, pp. 684–698, 2020.

[13] N. Turpault, R. Serizel, A. Parag Shah, and J. Salamon, "Sound event detection in domestic environments with weakly labeled data and soundscape synthesis," in *Workshop on Detection and Classification of Acoustic Scenes and Events*, October 2019.

[14] I. Nolasco, B. Ghani, S. Singh, E. Vidaña-Vila, H. Whitehead, E. Grout, M. Emmerson, I. Kiskin, F. Jensen, J. Morford, A. Strandburg-Peshkin, L. Gill, H. Pamuła, V. Lostanlen, and D. Stowell, "Few-shot bioacoustic event detection at the DCASE 2023 challenge," in *Proc. of the 8th Detection and Classification of Acoustic Scenes and Events 2023 Workshop (DCASE2023)*, September 2023, pp. 146–150.

[15] X. Mei, X. Liu, M. D. Plumbley, and W. Wang, "Automated audio captioning: An overview of recent progress and new challenges," *EURASIP J. Audio Speech Music Process.*, vol. 2022, no. 1, oct 2022.

[16] H. Xie, S. Lipping, and T. Virtanen, "Language-based audio retrieval task in DCASE 2022 challenge," in *Proc. of Detection and Classification of Acoustic Scenes and Events 2022 Workshop (DCASE2022)*, 2022.

[17] K. Choi, J. Im, L. M. Heller, B. McFee, K. Imoto, Y. Okamoto, M. Lagrange, and S. Takamichi, "Foley sound synthesis at the DCASE 2023 challenge," in *Proc. of the 8th Detection and Classification of Acoustic Scenes and Events 2023 Workshop (DCASE2023)*, September 2023, pp. 16–20.

[18] D. Barchiesi, D. Giannoulis, D. Stowell, and M. D. Plumbley, "Acoustic scene classification: Classifying environments from the sounds they produce," *IEEE Signal Process. Mag.*, vol. 32, no. 3, pp. 16–34, 2015.

[19] A. Mesaros, T. Heittola, E. Benetos, P. Foster, M. Lagrange, T. Virtanen, and M. D. Plumbley, "Detection and classification of acoustic scenes and events: Outcome of the DCASE 2016 challenge," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 26, no. 2, pp. 379–393, 2017.

[20] A. Mesaros, T. Heittola, and T. Virtanen, "A multi-device dataset for urban acoustic scene classification," in *Proc. of the Detection and Classification of Acoustic Scenes and Events 2018 Workshop (DCASE2018)*, November 2018, pp. 9–13.

[21] T. Heittola, A. Mesaros, and T. Virtanen, "Acoustic scene classification in DCASE 2020 challenge: Generalization across devices and low complexity solutions," in *Proc. of the Detection and Classification of Acoustic Scenes and Events 2020 Workshop (DCASE2020)*, November 2020, pp. 56–60.

[22] I. Martín-Morató, F. Paissan, A. Ancilotto, T. Heittola, A. Mesaros, E. Farella, A. Brutti, and T. Virtanen, "Low-complexity acoustic scene classification in DCASE 2022 challenge," in *Proc. of the 7th Detection and Classification of Acoustic Scenes and Events 2022 Workshop (DCASE2022)*, November 2022.

[23] F. Schmid, P. Primus, T. Heittola, A. Mesaros, I. Martín-Morató, K. Koutini, and G. Widmer, "Data-efficient low-complexity acoustic scene classification in the DCASE 2024 challenge," in *Proc. of the 9th Detection and Classification of Acoustic Scenes and Events 2024 Workshop (DCASE2024)*, 2024.

[24] A. Mesaros, A. Diment, B. Elizalde, T. Heittola, E. Vincent, B. Raj, and T. Virtanen, "Sound event detection in the DCASE 2017 challenge," *IEEE/ACM Trans. Audio Speech Lang. Process.*, 2019.

[25] N. Turpault, S. Wisdom, H. Erdogan, J. R. Hershey, R. Serizel, E. Fonseca, P. Seetharaman, and J. Salamon, "Improving sound event detection in domestic environments using sound separation," in *Proceedings of the Detection and Classification of Acoustic Scenes and Events 2020 Workshop (DCASE2020)*, November 2020, pp. 205–209.

[26] F. Ronchini and R. Serizel, "Performance and energy balance: A comprehensive study of state-of-the-art sound event detection systems," in *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2024, pp. 1096–1100.

[27] I. Martín-Morató and A. Mesaros, "Strong labeling of sound events using crowdsourced weak labels and annotator competence estimation," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 31, pp. 902–914, 2023.

[28] S. Cornell, J. Ebbers, C. Douwes, I. Martín-Morató, M. Harju, A. Mesaros, and R. Serizel, "DCASE 2024 task 4: Sound event detection with heterogeneous data and missing labels," in *Proc. of the 9th Detection and Classification of Acoustic Scenes and Events 2024 Workshop (DCASE2024)*, 2024.

[29] S. Adavanne, A. Politis, J. Nikunen, and T. Virtanen, "Sound event localization and detection of overlapping sources using convolutional recurrent neural networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 1, pp. 34–48, March 2018.

[30] K. Shimada, A. Politis, P. Sudarsanam, D. A. Krause, K. Uchida, S. Adavanne, A. Hakala, Y. Koyama, N. Takahashi, S. Takahashi, T. Virtanen, and Y. Mitsufuji, "STARSS23: An audio-visual dataset of spatial recordings of real scenes with spatiotemporal annotations of sound events," in *Advances in Neural Information Processing Systems*, A. Oh, T. Neumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, Eds., vol. 36. Curran Associates, Inc., 2023, pp. 72 931–72 957.

[31] D. A. Krause, A. Politis, and A. Mesaros, "Sound event detection and localization with distance estimation," in *32nd European Signal Processing Conference 2024 (EUSIPCO 2024)*, 2024, pp. 286–290.

[32] J. Liang, I. Nolasco, B. Ghani, H. Phan, E. Benetos, and D. Stowell, "Mind the Domain Gap: A Systematic Analysis on Bioacoustic Sound Event Detection," 2024. [Online]. Available: https://arxiv.org/abs/2403.18638

[33] Y. Koizumi, Y. Kawaguchi, K. Imoto, T. Nakamura, Y. Nikaido, R. Tanabe, H. Purohit, K. Suefusa, T. Endo, M. Yasuda, and N. Harada, "Description and discussion on DCASE2020 challenge Task2: Unsupervised anomalous sound detection for machine condition monitoring," in *Proc. of the Detection and Classification of Acoustic Scenes and Events 2020 Workshop (DCASE2020)*, November 2020, pp. 81–85.

[34] Y. Kawaguchi, K. Imoto, Y. Koizumi, N. Harada, D. Niizumi, K. Dohi, R. Tanabe, H. Purohit, and T. Endo, "Description and discussion on DCASE 2021 challenge task 2: Unsupervised anomalous detection for machine condition monitoring under domain shifted conditions," in *Proc. of the 6th Detection and Classification of Acoustic Scenes and Events 2021 Workshop (DCASE2021)*, November 2021, pp. 186–190.

[35] T. Nishida, N. Harada, D. Niizumi, D. Albertini, R. Sannino, S. Pradolini, F. Augusti, K. Imoto, K. Dohi, H. Purohit, T. Endo, and Y. Kawaguchi, "Description and discussion on DCASE 2024 challenge task 2: First-shot unsupervised anomalous sound detection for machine condition monitoring," *In arXiv e-prints: 2406.07250*, 2024.

[36] K. Drossos, S. Adavanne, and T. Virtanen, "Automated audio captioning with recurrent neural networks," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, Oct. 2017.