forum acusticum 2026
8-12 September · Graz, Austria

SOUND SOURCE DOMINANCE IN MOUNTAIN SOUNDSCAPES: MACHINE LISTENING, HUMAN PERCEPTION, AND MEMORY

Sara Lenzi[1,2], Simone Spagnol[3], Tin Oberman[4], Angelos Tsaligopoulos[5], Simone Torresin[5]

[1] *Faculty of Engineering, University of Deusto, Bilbao, Spain*
[2] *Ikerbasque Basque Foundation for Science, Bilbao, Spain*
[3] *Department of Architecture and Arts, Iuav University of Venice, Italy*
[4] *Institute for Environmental Design and Engineering, University College London, United Kingdom*
[5] *Department of Civil, Environmental, and Mechanical Engineering, University of Trento, Italy*

This paper is a revised version of the authors' submitted manuscript that was peer-reviewed by at least two anonymous reviewers.

Abstract

This paper evaluates sound source dominance in Alpine mountain soundscapes by comparing three distinct methodologies: human perception in situ, automatic sound event detection (SED), and retrospective judgement. Findings indicate that while SED and human perception ratings are generally consistent for categories such as human sounds and water, retrospective recollections can be influenced by non-acoustical factors and prior knowledge. Notably, traffic noise is often overestimated in memory, while natural sounds like water and wildlife are retrospectively associated with an overall experience of silence. These results suggest that integrating quantitative and qualitative data is essential for a comprehensive understanding of mountain soundscapes to support effective environmental monitoring and protective measures.

Keywords: *Mountain soundscapes, machine listening, perceived sound source dominance, retrospective sound source identification*

1. Introduction

Soundscape perception of a place is largely determined by the sound sources present and the overall context in which they are experienced [1]. When asked to describe a soundscape, people most often refer to the collection of sounds they hear in a specific moment in time, rather than to their feelings [2]. In natural and mountainous settings, residents and visitors alike experience benefits from intellectual, spiritual and symbolic interactions with nature that go beyond purely recreational aspects, and are also driven by soundscape. In such environments, human-generated sounds are generally perceived as highly disruptive, whereas natural sounds, particularly those associated with water, are consistently linked to higher perceived pleasantness [3], [4], [5]. Identifying the dominant sound source is among the key pieces of information collected in Method A of ISO/TS 12913-2. However, the collection of subjective ratings can be labour-intensive and depends on the presence of participants who actively assess the soundscape in situ. To address these limitations, a substantial body of research has explored automatic sound event detection (SED) as a means of analysing environmental audio recordings [6]. Early SED systems relied on engineered acoustic features and shallow classifiers, but the field advanced significantly with the adoption of deep learning architectures, particularly convolutional recurrent neural networks (CRNNs), which became the dominant framework for polyphonic SED and have been widely used in the Detection and Classification of Acoustic Scenes and Events (DCASE) challenges [7]. More recently, research has increasingly shifted toward attention-based transformer models, which better capture long-range temporal dependencies and enable end-to-end event-level detection [8]. These developments, together with even more recent advances in large audio-language models (LALMs), are beginning to enable models that jointly understand sound events, semantic context, and affective content, suggesting new opportunities for scalable and automated soundscape analysis [9].

In this study, using binaural recordings collected during mountain soundwalks, we evaluated the effectiveness of the established lightweight sound event classifier YAMNet,[1] as its pre-trained embeddings provide a convenient proxy for identifying dominant sound sources in environmental recordings. Its outputs were compared with the source dominance reported by par-

***Corresponding author**: sara.lenzi@deusto.es.*

[1] `https://github.com/tensorflow/models/tree/master/research/audioset/yamnet`

ticipants during the soundwalks through an extended version of the questionnaire included in Method A of ISO/TS 12913-2, as well as with the sound sources recalled by the same participants in post-soundwalk interviews. The objectives were twofold, and aimed at providing a novel, triangulated approach to evaluating methodologies for assessing sound source dominance in mountain environments: first, to assess the performance of YAMNet in predicting dominant sound sources in natural mountain contexts; and second, to compare visitors' auditory judgments when immersed in such environments versus in post-hoc recall scenarios, where reporting may be influenced by different situated, emotionally driven responses.

1.1. Background and context of the research

The study presented in this paper was conducted between June and October 2025 in the framework of the project *Sounding Trentino* which aims at investigating the relationship between mass tourism, soundscape quality, and potential for protective measures in the Alpine Province of Trento, in Northern Italy. Four sites were selected that are both well-known touristic destinations as well as representative of the territory. The sites were all accessible, with well-defined hiking trails of low-to-medium difficulty. The sites shared similar socio-ecological character being classified by the territorial plan of the Autonomous Province of Trento[2] as Alpine areas marked by forestation.

In this paper, we focus on the assessment of dominant sound sources across the four sites. We discuss similarities and differences between automatic and human identification of dominant sound sources, and how they compare with retrospective soundscape judgements.

2. Materials and Methods

A soundwalk was conducted in each of the selected sites. Overall, 67 people took part in the activities, aged 14 to 75, 66% women, 70% practising mountain sports at least once a month. With the exception of four people, all participants were locals, i.e., permanent residents or native of the territory under study.

2.1. Audio data collection and questionnaire

Each soundwalk was organised in a number of intermediate stops, from a minimum of four to a maximum of five stops. At each stop, participants were instructed to listen to the surrounding soundscape for three minutes, in silence. Acoustic measurements and audio recordings were gathered by an operator wearing a BHS II head-mounted binaural microphone (HEAD Acoustics). Throughout the silent intervals, both the operator and the participants faced a uniform direction. A 1 kHz sine-wave generator (94 dB) was employed to calibrate the recording system prior to each experimental block. In all soundwalks, the first stop coincides with the parking lot at the beginning of the natural path, sometimes with a nearby bar. Similarly, the last stop coincides with the arrival at a mountain hut or restaurant, with the exception of one soundwalk (Viote). In between these two stops, all other stops happened along an official (i.e., clearly marked by local institutions) natural path. After the completion of the listening and recording activity, participants were asked to fill a version of the ISO/TS 12913-2 Method A (Annex C), adjusted to capture more details around natural sound types, as in [4].

2.2. Automatic sound event detection

Automatic SED was performed using YAMNet, a lightweight convolutional neural network developed by Google for general-purpose audio event classification [10]. YAMNet is trained on the AudioSet dataset, a large-scale collection of human-labelled audio clips drawn from YouTube and organized according to a hierarchical ontology comprising over 500 sound event classes [11]. The model operates on log-mel spectrogram patches extracted from the input signal and produces frame-level outputs consisting of a probability distribution over all AudioSet classes for each frame, from which class predictions are derived.

All binaural recordings collected during the soundwalks were processed using the pre-trained YAMNet model without further fine-tuning. Since manual frame-level annotation was not conducted on the recordings, local classification metrics cannot be calculated; however, the off-the-shelf YAMNet architecture carries an established mean Average Precision of 0.31 on the AudioSet corpus. In order to align the model output with the survey categories, a subset of AudioSet classes was mapped onto six macro-categories, i.e., *traffic*, *other noise*, *human sounds*, *animal sounds*, *wind*, and *water*, according to the semantic meaning of each class within the ontology. Importantly, AudioSet classes not relevant to the target taxonomy (e.g., music) were excluded from the analysis.

Since YAMNet is designed for mono input, each binaural recording was split into its left and right channels, which were processed independently. For each channel, frame-level predictions were obtained from YAMNet and converted into categorical labels by selecting the class with the highest posterior probability at each time frame. When the predicted class corresponded to one of the selected AudioSet classes, it was assigned to the corresponding macro-category; otherwise, the frame was treated as unclassified. The temporal presence of each macro-category was then quantified by computing the proportion of frames in which that category was detected relative to the total number of frames in the recording.

Finally, to obtain a single descriptor per recording while preserving the most prominent information captured binaurally, the percentage values computed for the left and right channels were compared, and for each macro-category the maximum value between the two channels was retained. This resulted in a set of six percentages for each audio clip, representing the relative temporal dominance of each sound source

[2] http://www.territorio.provincia.tn.it/portal/server.pt/community/piano_urbanistico_provinciale/768/piano_urbanistico_provinciale/21168

Table 1: Median perceived dominance ratings (P, as 5-point Likert scales) assigned by participants to each of the six sound source categories (traffic, other noise, human sounds, animal sounds, wind, and water) and relative temporal presence of the same categories obtained through automatic detection (D, as %).

Site	Stop	Traffic		Other		Human		Animal		Wind		Water	
		P	D	P	D	P	D	P	D	P	D	P	D
Nambino	1	4	9.00	3.5	14.54	3.5	42.45	1	4.82	3	0	1	0
	2	2	0	1	0	4	75.57	2	1.68	1.5	0	1	0
	3	1	0.32	1	0.26	3.5	47.91	2	3.14	3.5	0	3.5	0.06
	4	1	2.52	1	0	3	17.70	2	0	2	0	5	77.84
	5	1	1.26	1	0	2.5	23.37	3	4.59	3	0	4.5	0.53
Cimana	1	4	18.68	1	1.23	4	49.15	3	3.82	1	0	1	0
	2	2	0	1	0	2	2.37	3	19.53	2	0	1	0
	3	2	0	1	0	2	0.82	3	29.70	1.5	0	1	0
	4	2	0	1	0.62	4	32.79	3	10.10	1	0	1	0
	5	1	0	1	0	4	94.45	3	0	1	0	1	0
Cembra	1	4	35.51	1	5.04	1	0.72	3	0.59	1	0	1	0
	2	3	0.26	1	1.66	1	0.46	3.5	0.79	3	0	1	0.33
	3	3.5	5.67	1	0	3	0.86	3	0.20	2	0.33	1	0
	4	3	0	2	0.87	1	0.31	2	2.67	3	0	1	0
	5	3	4.44	2	0.19	1	0	2	0.26	4	12.10	1	0.45
Viote	1	2	0.40	1	0	4	49.39	3	3.23	2	0	1	0
	2	1	0	3	0	3	5.55	3	4.03	1	0	1	0
	3	2	0.19	1	0	3	19.52	2	3.02	3	0	1	0.13
	4	1	0	2	0	2	0.40	4	13.11	3	0	1	0

category as derived from the automatic analysis.

2.3. Post-soundwalk interviews

Semi-structured interviews were conducted in person with selected participants (n=10, 6 women, 4 men) about two weeks after the respective soundwalk, and lasted for about 40 minutes. All interviewees were local, i.e., permanent residents of the places under study who acted as context-specific experts of the local soundscape, capable of identifying and interpreting sound sources based on lived experience. This responded to a specific choice of the *Sounding Trentino* project, which is funded by local institutions with the aim of assessing the impact of tourism on the Alpine soundscape *as experienced by local communities*. Interviews were recorded, transcribed, and coded using the Atlas.ti[3] software, following an iterative thematic analysis methodology [12]. Through open-ended questions, three main themes were explored: the relationship of the interviewee with the site; opinions and feelings toward mountain tourism; sound events recalled as being heard during the walk. In this paper, we only report insights from the last question.

3. Results

3.1. Perceived dominant sound sources

Table 1 reports the median ratings (5-point Likert scale) assigned by participants to the dominance of each sound source category (traffic, other noise, human sounds, animal sounds, wind, and water) at each listening stop. Although some variability was observed among listening points within each site, a site-level overview reveals clear patterns. Cembra is strongly characterized by traffic noise, which was rated as moderately dominant or higher across all listening stops, while it is nearly absent in Viote. Human sounds are generally prominent across these tourist sites, with the exception of Cembra. Animal sounds are perceived at all sites, with higher prevalence in Cimana and lower dominance in Nambino. Wind was more pronounced in Nambino and Cembra, whereas it was barely perceptible in Cimana. Water sounds are largely absent across sites, except for Nambino, where the presence of the lake contributes substantially.

3.2. Comparison of automatic detection and perceived sound dominance

For each site and listening point, the relative temporal presence of the six sound source categories obtained through SED was compared with the perceived dominance reported by participants via correlation analysis. This approach allows evaluating the extent to which the temporal presence of a given sound category as identified by YAMNet is associated with its perceived dominance in situ. Spearman's correlation coefficients were computed by matching the mean in-situ dominance ratings (from 5-point Likert scales) with the relative temporal presence (% of total clip duration) of each AudioSet macro-category across all stops.

The results show a generally positive relationship between automatic detection and perceptual evaluation across all categories, although with notable differences

[3] https://atlasti.com

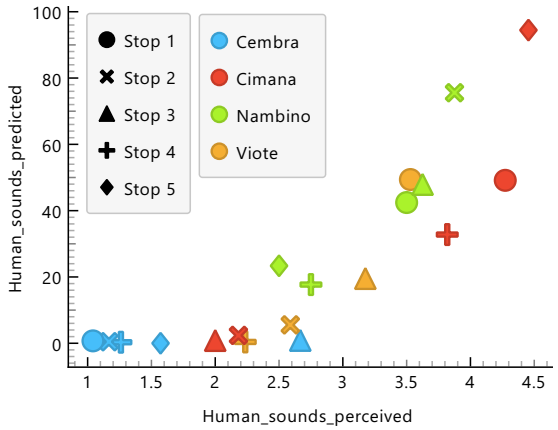

Figure 1: Correlation between relative temporal presence (%) as given by YAMNet and average perceived dominance of the *human sounds* category.

in strength. As shown in Figure 1, the highest correlation was observed for *human sounds* ($\rho = .911$, $p < .001$), followed by *water* ($\rho = .650$, $p = .003$), suggesting that these sound events are consistently identified both algorithmically and perceptually. Moderate correlations were found for *other noise* ($\rho = .459$, $p = .048$) and *traffic* ($\rho = .434$, $p = .063$), likely reflecting the less clearly defined acoustic characteristics of these types of sounds. In contrast, *animal sounds* ($\rho = .275$, $p = .255$) and *wind* ($\rho = .267$, $p = .270$) showed notably weak correlations, possibly reflecting the subtle nature of these sounds which can be faint (e.g., distant bird vocalizations) and, in the case of wind, partly influenced by the use of wind shields on the microphones.

3.3. Retrospective sound event identification: Insights from semi-structured interviews

Figure 2 recaps on the different sound events mentioned across all sites, and the relative proportion of each. In addition to higher-level categories aligned with the questionnaires and the SED (i.e., human sounds, animal sounds, traffic, water, and wind), two categories emerge strongly that are not identified through the other two methods: *noise* and *silence*.

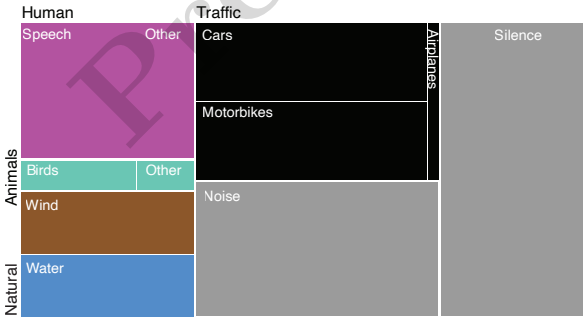

Figure 2: Sound events recalled by interviewees, clustered by higher level categories. The treemap diagram shows the relative weight of each sub-category across all sites.

When prompted to reflect on the sounds heard during the walk, *silence* is in fact the most common recollection for all soundwalk sites, with the exception of Cembra. Comments such as *'I was surprised by the absence of sounds'*, *'We were in a sort of bubble of silence'*, speak of the perception of a peaceful, secluded environment, where *'Someone who wants to be alone with one's thoughts is in the right place'*.
As for *noise*, this differs from the *other noise* category in the questionnaire and the SED. If in the latter, noise includes sounds as sirens, construction, industry, loading of goods, in the interviews the word noise is used to reinforce other sound events associated with negative feelings, either from traffic or human sounds: *'At first there were only car sounds, and all that noise did not make sense in that Alpine environment'*.

3.4. Integrated analysis

Figure 3 shows sound events as they were retrospectively identified for each individual site, and the number of occurrences for each category. *Silence* is reported (although it is not, strictly speaking, a sound event) due to its relevance in the interviewees' recollection as described in the previous section.

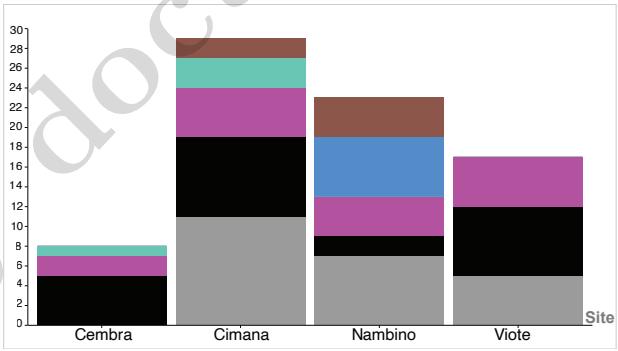

Figure 3: Sound event categories as recalled by interviewees for each individual site and number of occurrences. *Human sounds* in pink, *traffic* in black, *animal sounds* in turquoise, *water* in blue, *wind* in brown, *silence* in grey.

As shown in Figure 3, *water* is recalled as a dominant sound source for the Nambino site. This is consistent with the findings from questionnaires and the SED (see Table 1), where water is reported from moderate to completely dominant for stops 3 to 5. The path to the Nambino Lake, in fact, is characterised by the presence of a creek, specifically in stops 3 and 4. *Wind* is also reported as moderately dominant for the same stops in the questionnaires, and confirmed in the interviews: *'[...] we had the wind behind that you could see the plants moving, and the waterfall at the back – it was the energy of nature, wind and water'*.
Interestingly, *human sounds*, which are reported as moderately to completely dominant in both questionnaire and SED except for Cembra, partially lose relevance in the recollection, with the notable exception of the restaurant area of Cimana (*'There were so much noise from people, I didn't expect that at all'*). Other occurrences of human speech in Nambino, Cembra,

and Viote refer to mostly families and children passing by, and are characterised by positive feelings (*'Those children's voices, one may indeed call it shouting, but in that context they were beautiful'*).

As for *traffic* sounds (motorbikes and cars), these sources acquire particular importance, especially in Cimana and Viote, where they are also consistently associated with *noise*. While for Cimana *traffic* is clearly situated at the first and last spots (*'The walk has been a long relaxing silence in between two traffic noise spots'*), thus aligning with the questionnaires and the SED, for Viote the latter do not report high traffic. In Cembra, the dominance of *traffic* sounds is clear (see Figure 3) and consistent across all the three analysis methods (*'I was shocked by how much you can hear the traffic sounds coming up the valley'*). Notably, while in Cembra the questionnaires and the SED point at *wind* as another clearly identifiable sound source, no interviewees mentioned it in their recollection.

Lastly, *animal sounds* are in general less recalled in the interviews than in the questionnaire and SED, across all sites. While in Cimana interviewees only mention remembering *'a bird here and there'* within a general context of silence (see Section 3.3), results from the questionnaire from the same site report a moderate dominance of animal sounds across all stops, echoed by the SED, especially in the central spots of the walk (e.g., into the woods). Notably, in Viote *animal sounds* are never mentioned, while they are moderately to highly dominant according to the questionnaire and the SED, especially in the last stop (see Table 1). In Cembra, *animal sounds* are never mentioned by interviewees (nor obtained high dominance rating through the SED), while they are perceived as moderate to high-moderate dominance in the questionnaire.

4. Discussion

When compared with the retrospective judgement, some categories stand out as being consistent – although to different degrees – with the other two methods, notably *water* and *human sounds* (see Section 3.4). *Traffic* sounds are, conversely, generally overestimated in the retrospective assessment, and often mentioned as a negative or unwanted component of the soundscape. At this stage, it is unclear to what extent such comments reflect consolidated prior experiences. Recall that interviewees are local, therefore, comments such as *'They keep the engine of the bike on all the time, seems like they only want to get some noise heard'* might reflect their general perception of the site, rather than events identified during the walk. On the contrary, *animal sounds* and *human sounds* were in general under-evaluated in the retrospective assessment, especially for Nambino. This is particularly striking as Nambino is a well-known destination for non-local tourists, perhaps the most famous of the tourist sites chosen for this study. Other comments reflect the relationship with wildlife, with human sounds being appreciated as a defence against the presence of the bear (a very well-known and controversial topic in the region): *'I'd rather hear other people's voice these days, you know, for the bear'*.

These findings are aligned with Jo and Jeon [13] who, expanding on [14], observed that when comparing sound event dominance through ISO 12913-2 Method A (questionnaires) and Method C (retrospective interviews), results from the latter are influenced by non-acoustical factors, such as prior knowledge of the site. In future development of this study, it would be valuable to investigate how the acoustic memories of the sites were influenced not only by prior experience but also by expectations generated by, e.g., public controversies (over-tourism, conflict with wildlife) or exposure to idealised images (e.g., on social media). Differentiating such influence between local residents and tourists would also be valuable.

A separate consideration should be given to the importance of *silence* as the most commonly mentioned 'sound' across all sites (Figure 2). It is worth noticing that the perception of silence is situated in the sites where natural sounds are dominant – particularly *water* (in Nambino) or *animal sounds* (in Cimana and Viote). In Cembra, where all the analysis methods point at traffic sounds as dominant, silence is never mentioned in the interviews. This might suggest that, in these contexts, *silence* is not retrospectively experienced as the absence of sound, but rather as the presence of sounds that are coherent with the environment (e.g., water and wildlife), provided they are not disrupted by traffic noise. It is worth noticing that for the same sites (i.e., all sites with the exception of Cembra), interviewees consistently report feelings of tranquillity and peace, also connected with the concept of silence. This interpretation would confirm findings from [15], where personal and situational factors (in this case, the overall experience of peace during the walks) significantly contributed to retrospective judgements of soundscapes. The absence of silence in the other analysis methods is however expectable, since (1) silence is not among the sound sources listed in the questionnaire, and (2) although silence is included among the classes of the AudioSet ontology, its detection is dramatically dependent on input level, making it unreliable in the present context.

Finally, certain methodological limitations of the SED approach should be acknowledged. First, deploying YAMNet without localized domain adaptation may under-detect low-amplitude, subtle biophonic or geophonic events due to its web-video training baseline. Second, the manual mapping of AudioSet classes into broader categories is inherently subjective and could introduce classification biases, though this aggregation was required to allow a direct comparison with the perceptual survey data.

5. Conclusions

In this study, we used three different methods for the assessment of sound dominance in four Alpine locations. In general, positive trends can be observed between dominant sound sources identified by par-

ticipants to the soundwalk (via the ISO/TS 12913-2 Method A), automatic classification of sound events (via YAMNet), and the post-experience interviews (Method C), especially for the categories of *water* and *human sounds*. However, this correspondence is less consistent for *traffic* sounds, which are remembered as more prominent than actually experienced, and seem to be, by far, the greatest source of unwanted sounds in the Alpine context. Interestingly, at sites less affected by traffic where water and animal sounds were reported, these elements are retrospectively associated with the experience of silence.

Moving forward, the integration of quantitative and qualitative data, and the comparison of local listeners' judgements with tourists could provide a comprehensive understanding of the Alpine soundscape in support of monitoring and protective measures that would facilitate the cohabitation of residents, tourists, and wildlife in mountain areas.

6. Acknowledgments

The project *Sounding Trentino* is funded by the *Visiting in Trentino* 2024 programme of the Provincia Autonoma di Trento, Italy. SL work is partially funded by the Marie Skłodowska-Curie Cofund Programme of the European Commission (H2020-MSCA-COFUND-2020-101034228-WOLFRAM2). TO work is funded by HEAD Genuit Stiftung (Grant no 25/04-W). We wish to thank Fondazione Cassa Rurale di Trento for the organisation of Cembra and Viote soundwalks.

References

[1] Ö. Axelsson, M. E. Nilsson, and B. Berglund, "A principal components model of soundscape perception," *Journal of the Acoustical Society of America*, vol. 128, no. 5, pp. 2836–2846, 2010.

[2] G. Gozzi, S. Torresin, and L. Badan, "Soundscapes across mountains and cities: A linguistic study in the Trentino region," *Acoustics*, vol. 8, no. 1, 2026.

[3] R. Ferrari, R. Rupf, and B. Reutz, "Alpine soundscapes: Sounds and their consequences for perceived recreational quality – A case study of two Regional Nature Parks – Beverin Nature Park and Parc Ela in Switzerland," *Eco Mont – Journal on Protected Mountain Areas Research & Management*, vol. 15, no. 2, pp. 4–12, 2023.

[4] T. Oberman, A. Latini, F. Aletta, G. Gozzi, J. Kang, and S. Torresin, "Human sounds and associated tonality disrupting perceived soundscapes in protected natural areas," *Scientific Reports*, vol. 15, no. 1, 2025.

[5] M. Ebner and U. Schirpke, "Participatory mapping of human-nature interactions in mountain landscapes: Instances of experiential multifunctionality," *Landscape and Urban Planning*, vol. 263, no. 105453, 2025.

[6] A. Mesaros, T. Heittola, T. Virtanen, and M. D. Plumbley, "Sound event detection: A tutorial," *IEEE Signal Processing Magazine*, vol. 38, no. 5, pp. 67–83, 2021.

[7] A. Mesaros, R. Serizel, T. Heittola, T. Virtanen, and M. D. Plumbley, "A decade of DCASE: Achievements, practices, evaluations and future challenges," in *Proc. of the 2025 IEEE International Conference on Acoustics, Speech and Signal Processing*, Hyderabad, India, 2025.

[8] K. Zaman, K. Li, M. Sah, C. Direkoglu, S. Okada, and M. Unoki, "Transformers and audio detection tasks: An overview," *Digital Signal Processing*, vol. 158, no. 104956, 2025.

[9] H. Yin and J.-W. Choi, *Can Large Audio Language Models understand audio well? Speech, scene and events understanding benchmark for LALMs*, 2026. arXiv: `2509.13148 [cs.SD]`.

[10] S. Hershey et al., "CNN architectures for large-scale audio classification," in *Proc. of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing*, New Orleans, USA, 2017, pp. 131–135.

[11] J. F. Gemmeke et al., "Audio Set: An ontology and human-labeled dataset for audio events," in *Proc. of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing*, New Orleans, USA, 2017, pp. 776–780.

[12] V. Braun and V. Clarke, "Thematic analysis," in *APA handbook of research methods in psychology, Vol. 2. Research designs: Quantitative, qualitative, neuropsychological, and biological*, H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, and K. J. Sher, Eds., Washington, DC, USA: American Psychological Association, 2012, pp. 57–71.

[13] H. I. Jo and J. Y. Jeon, "Comparing soundscape assessment methods of ISO 12913-2 with questionnaires (Methods A and B) and narrative interview (Method C)," in *Proc. of Forum Acusticum*, Lyon, France, 2020, pp. 1449–1452.

[14] F. Aletta, C. Guattari, L. Evangelisti, F. Asdrubali, T. Oberman, and J. Kang, "Exploring the compatibility of Method A and Method B data collection protocols reported in the ISO/TS 12913-2:2018 for urban soundscape via a soundwalk," *Applied Acoustics*, vol. 155, pp. 190–203, 2019.

[15] J. Steffens, D. Steele, and C. Guastavino, "Situational and person-related factors influencing momentary and retrospective soundscape evaluations in day-to-day life," *Journal of the Acoustical Society of America*, vol. 141, no. 3, pp. 1414–1425, 2017.