What in the world do we hear? Understanding public and private spaces through SoundAI

Sara Lenzi
Faculty of Engineering, University of Deusto, Bilbao, Spain; bIkerbasque Basque Foundation for Science, Bilbao, Spain

Elif Özcan
Critical Alarms Lab, Delft University of Technology, Delft, The Netherlands

Simone Mora
Sensable City Lab, Massachusetts Institute of Technology, Cambridge, United States of America

Martina Mazzarello
Sensable City Lab, Massachusetts Institute of Technology, Cambridge, United States of America

Aase Haatveit
Sensable City Lab, Massachusetts Institute of Technology, Cambridge, United States of America

See next page for additional authors

Citation

Lenzi, S., Özcan, E., Mora, S., Mazzarello, M., Haatveit, A., and Duarte, F. (2024) What in the world do we hear? Understanding public and private spaces through SoundAI, in Gray, C., Hekkert, P., Forlano, L., Ciuccarelli, P. (eds.), *DRS2024: Boston*, 23–28 June, Boston, USA. https://doi.org/10.21606/drs.2024.1550

Authors
Sara Lenzi, Elif Özcan, Simone Mora, Martina Mazzarello, Aase Haatveit, and Fabio Duarte

What in the world do we hear? Understanding public and private spaces through SoundAI

Sara Lenzi[a,b,c]*, Elif Özcan[c], Simore Mora[d,e], Martina Mazzarello[d], Aase Haatveit[d], Fabio Duarte[d]

[a]Faculty of Engineering, University of Deusto, Bilbao, Spain
[b]Ikerbasque Basque Foundation for Science, Bilbao, Spain
[c]Critical Alarms Lab, Delft University of Technology, Delft, the Netherlands
[d]MIT Senseable City Lab, Massachusetts Institute of Technology, Cambridge, United States of America
[e]Department of Computer Science, Norwegian University of Science and Technology, Oslo, Norway

*Corresponding e-mail: sara.lenzi@deusto.es

doi.org/10.21606/drs.2024.1550

Abstract: This paper reports on the Lab that the authors conducted during the DRS2024 conference. Through novel concepts of "Auditory Footprint," the authors aimed at exploring and highlighting the perceptual and ecological dimensions of sound events in both private and public spaces. During the DRS2024 Lab, multidisciplinary experts explored AI's capacity to interpret and influence human interactions with soundscapes, proposing speculative-use cases, and addressing challenges such as subjective perception, equitable data collection, and ethical considerations. The findings advocate for a responsible integration of SoundAI technologies, aiming to foster accessibility, environmental awareness, and community well-being, while mitigating risks. Future work will formalize frameworks for AI-based sound applications and explore policy and technical feasibility.

Keywords: soundscape; Machine Listening; Artificial Intelligence; human-data interaction

1. Introduction

Our living environments, both private and public spaces, have changed over time and can impact many elements of our personal, social, and professional lives, both in the individual and collective dimensions. Researchers, designers, and representatives of various industries are increasingly interested in intuiting and forecasting how people are impacted by their surroundings. People change the way they interact with the spaces they inhabit, and shape the environments according to the actions they have to perform. Acquiring and analysing data of human behaviour and ways of living have been traditionally very challenging due to the private nature of the home environment and the complexity and diversity of living contexts around the world. Today, the growing presence of technology in our lives (i.e., intelligent devices and IoT at home) and their seamless integration with our daily actions – including new habits (e.g., quantifying our activities and creating digital twins of our body and thoughts), enrich the possibility of using opportunistic data to reveal certain aspects of our lives. In the realm of sound, for example, smart speakers and music services have brought the idea of AI-assisted sound analysis and reproduction to the broader market. However, we are still far from using SoundAI research (e.g., research on generative AI for music or sound effects production, and automated identification of sound events through Machine Learning) to impact people's sound-conscious behaviour in their day-to-day lives. This is due to several issues, such as awareness of concerns about privacy, and unclear practical applications or commercial opportunities.

At the same time, there is growing interest in the study and improvement of our acoustic environments. The soundscape approach (i.e., studying the acoustic environment as perceived and experienced by listeners) is becoming a common ground of research with interesting application domains emerging, including analysing sound in urban spaces such as building facades, sidewalks, streets (Callero et al., 2018; Bello et al., 2019) also as an indicator of social behaviour (Lenzi et al., 2021), and computationally modelling the changes in outdoor Sound Pressure Levels (Caniato et a., 2021; Salazar Miranda et al., 2020). Computational models of the acoustic environment have been developed for analysing indoor soundscapes (Ohran, 2019; Spagnol et al., 2022) with the goal of comparing people's qualitative perception of sounds (Torresin et al., 2021; Torresin et al., 2019) or of analysing and forecasting the impact of noise on human cognition (Mehta et al., 2012) and on environments that are shared (e.g. urban space, hospitals, for a recent review, see Lenzi et al., 2023).

However, these early approaches miss the possibility of acquiring insights into how Artificial Intelligence – a growing field of research in, for instance, vision-related fields (e.g., image recognition or generation) – interprets the traces of sound events, understands the activities that produced them, and measures their impact on the shared environment. At the moment, the field of computational sound analysis is still lacking research conducted in living environments (indoors and outdoors), a necessary step to explore which are the possibilities of an 'AI for sound', its future application scenarios, and its impact on the lives of humans and other species.

1.1 Auditory footprints: On the impact of sound in everyday life

The ability to construct meaning from listening to the world is central to human perception. As humans, we rely on sound events to make sense of, interact with, and sometimes predict, the behaviour of our surroundings in natural as well as artificial (urban, virtual) personal and shared environments (Lenzi et al., 2024). Recent research underscores the role of sound in fostering emotional connections between humans and robots (Liu et al., 2023). Empirical research highlights medical personnel's reliance on auditory cues in orthopaedic operating rooms to execute precise, sound-triggered actions (Özcan et al., 2022). Moreover, the COVID-19 pandemic highlighted the profound impact of altered soundscapes on human-environment interactions (Lenzi et al., 2021), interspecies relationships (Duarte et al., 2021), and even on plant life (Phillips et al., 2021). In the realm of cinema, the semantic, causal, and material dimensions of sound are employed to construct narratives and direct audience interpretation (Chion, 1994).

In its seminal work '*What in the world do we hear? An ecological approach to auditory event perception'* (1993) which served an inspiration for this Lab, Gaver introduces the definition of '*everyday sounds*':

> "Imagine that you are walking along a road at night when you hear a sound. [...] As you stand there in the road, it is likely that you will not listen to the sound itself at all. Instead, you are likely to notice that the sound is made by an automobile with a large and powerful engine. Your attention is likely to be drawn to the fact that it is approaching quickly from behind. And you might even attend to the environment, hearing that the road you are on is actually a narrow alley, with echoing walls on each side. This is an example of everyday listening, the experience of listening to events rather than sounds."

As we flow into countless different contexts, events, and interactions in our daily lives, sound - and the act of listening - help us make sense of the world around us by decoding and interpreting the information that the sound events carry (in Gaver's words, *"that the sound is made by an automobile with a large and powerful engine"*). Additionally, and inextricably, sounds also carry perceptual qualities to which we, as humans, learn to assign emotional and psychological values: the sound of an engine can be perceived as pleasant or unpleasant and even adventurous, depending on the context and our personal interests, beliefs, or prior experiences (Özcan, 2014). The soundscape of a public space can be perceived as chaotic or calm depending on our expectations, goals, or personal circumstances.

In our previous work (Lenzi et al., 2023), we coined the term *Auditory Footprint* to define a novel approach to the assessment of the quality of our everyday (private or public) soundscape, which combines the identification of the sound source, its physical properties, and its affective qualities as perceived by listeners. The word '*Auditory*' refers to the sense of hearing and here is used to capture the perceptual impact of sound on the hearing system. '*Footprint*' refers to those characteristics that help identify a particular sound event e.g., the source that produces it and its acoustic properties (frequency, amplitudes, timbre….). The Auditory Footprint of a sound in computational language would then aim to assess the perceived quality of a specific sound event through an aggregated index which - similarly to the Carbon Footprint - provides combined information on the impact of specific sound events on the environment as perceived by humans. Such a novel approach would leverage algorithmic modelling to inform end-users with the goals of increasing awareness of the role of sound in everyday life and of supporting better decision-making to mitigate the negative effects of sound pollution. Monitoring such footprints will improve the quality of the acoustic environment in spaces where issues of poor sound quality and its negative impact on the resident populations are well known - such as hospitals, urban spaces dominated by traffic and other noise sources, schools and other educational buildings, and crowded residential buildings (Lenzi et al., 2023).

But how would the Auditory Footprint of sound be measured, and what is the potential of Artificial Intelligence - and specifically, of Machine Listening (from now on, MLi i.e., Machine Learning techniques applied to the automatic identification and classification of sound events, Salomon and Bello et al., 2017) – for the automatic assessment of soundscape quality in public and private spaces?

1.2 Frontiers in SoundAI: Soundscape studies and machine learning for sound-based activity inference

All sound events carry information relevant to us and represent meaning (Özcan et al, 2012). Typically, the relation between humans and sound events is measured by the sound's physical power and its psychological effect on listeners (i.e., loud sounds may be arousing and soft and dull sounds more relaxing). Therefore, acoustic and psychoacoustic properties (e.g., sound's loudness over time, or the annoyance it causes to human listeners) are used as means of measurement. Bioacoustics (Sueur and Farina, 2015) and ecoacoustics (Farina et al., 2024) are tasked with assessing such characteristics in relation to non-human living creatures and their respective ecosystems (e.g., how is underwater life affected by marine industries?). Growing in traction, soundscape studies have been developing methods and techniques for the assessment of the affective quality of sound events (e.g., pleasantness and eventfulness) as perceived by humans in context (ISO 12913-1:2014).

In the most widely used soundscape evaluation protocol (Axelsson et al., 2010 and ISO 12913-3:2019), the affective qualities of sounds are measured through eight attributes, i.e., Calm, Pleasant, Exciting, Eventful, Chaotic, Annoying, Monotonous and Uneventful which are later combined in a predictive model of a given soundscape's quality (Mitchell, 2022). Additionally, to accurately describe how individual sounds (e.g., sounds of humans, alarms, machinery) contribute to the perception of such quality, sound sources are identified and classified manually by human listeners. Currently, meaning extraction for detected sound events is a retrospective analysis tasked on actual listeners as AI-based technologies are lagging behind (for instance, compared to the automated image recognition technology widely available even in consumer products).

Research that explores MLi to automatically classify sound events and computationally model the affective response of humans is on the rise. However, such research mainly focuses on outdoor environments (Lenzi et al., 2023; Merino-Martinez et al., 2021; Mitchell et al., 2021; Pedersen et al., 2018; Salamon et al., 2016; Broccolini et al., 2012). This is due to several concurring factors, among which is the presence of international legislation that requires local authorities to periodically monitor and map the noise levels in urban spaces (European Parliament, 2000), and the increased diffusion in cities of fixed sensor networks for the automatic collection of large volumes of soundscape data (Sevillano et al. 2016).

While there is clearly a scientific gap for intelligent solutions for the algorithmic modelling of the soundscape, a series of unresolved issues are faced, including critical ethical and privacy concerns in the collection of the required audio data for the training of algorithms, as well as in the application of such 'monitoring' technology. These concerns are recognized by providers of commercial solutions for audio data collection for AI (Javahid, 2023) as well as by researchers (Nautsch et al., 2019). Ideally, as a first step, a responsible and ethical AI for sound would support traditional sound measurement methods, and develop and incorporate computational models for sound source identification to provide more relevant information to the public (Lenzi et al, 2023).

As researchers interested in the monitoring of the soundscape quality with a goal towards its improvement, we engaged in a joint effort to explore the boundaries of the application of AI technologies for the automatic assessment of our everyday soundscapes and the potential impact (both positive and negative) of its development and application. The laboratory presented in this article is a first outcome of this effort.

2. DRS2024 Lab: SoundAI at the intersection of private and public spaces

The Lab we conducted during the 2024 DRS conference was part of an ongoing collaboration between the MIT Senseable City Lab (SCL) and the Critical Alarms Lab (CAL) at the Delft University of Technology (TU Delft). The collaboration is exploring the potential of MLi on our relationship with sound and the soundscape, and in particular the role and ethical implications of such technologies. The collaboration is taking place over three distinct phases, of which the Lab proposed for DRS2024 was the central phase.

Phase 1: Exploration

In an exploratory phase, we framed the guiding questions of the research, which aimed to investigate what AI-based agents can understand of human activities by listening to them, both in private and public spaces. Specific research questions included:

- What is the predictability of human activity based on a distributed microphone network?
- Can we detect how people inhabit and interact with their environments by analysing their soundscape?
- Can AI recognize and monitor the activity we perform in the privacy of our homes by listening to it?
- What are their possibilities and future scenarios in the context of private and public environments?

Phase 2: Speculative design session

During the DRS2024 conference, we conducted a one-day activity to engage domain experts such as sound experts, AI experts, design researchers, sociologists, other scholars, and activists, to address the theoretical, practical and ethical consequences of MLi applications when introduced into the real world. A detailed description of the Lab's setup and outcomes will follow.

Phase 3: Action list

The third phase of the Lab, currently ongoing, is based on the exploration of specific-use cases as they emerged from the DRS Lab. Current topics. Early topics include the automatic detection and generative design of soundscapes related to cultural identity in the context of migrations, and automatic detection of biodiversity and multi-species auditory sensitivity in the urban space through sound.

Figure 1. The setup of the Lab with three expert groups discussing the topics proposed and a view of the DRS design researchers consolidating the experts' final opinions.

During the DRS2024 conference, we specifically selected and invited experts to engage in a one-day activity to address the theoretical, practical, and ethical consequences of SoundAI applications (starting from, but not limited to, Machine Listening) when introduced into the real world. Nine experts from areas including sound computing and sonification, sound-driven design, acoustics, urban studies, public health, design, and performing arts joined the Lab for a closed-door half-day session. The second part of the Lab was devoted to an open-door session where experts shared the outcomes with a broader audience of DRS conference attendees with approximately 50 design researchers. The setup can be seen in Figure 1.

At the inception of the Lab, participants were split into three groups of three people, each assisted by a facilitator from the organisers' team. Experts were grouped according to their expertise with the explicit goal of fostering multi- and cross-disciplinary discussion (e.g., sound experts, who were the larger group, were equally split between each different group). Four ‘trigger questions’ – which were intended as a facilitator for discussion and had been identified during the Phase 1 of our collaborative research – were presented to the participants. The questions, listed below, were intended to provide initial hints for debate:

1. What is an ‘Auditory Footprint’? How does sound impact our relationship with the world? What should an algorithm that measures the Auditory Footprint of sound tell us that can empower users to improve the sonic environment?
2. What could AI for sound (and specifically, Machine Listening) be used for in a way that benefits society?
3. What happens at the interface between indoor sounds and the city?
4. What could AI for sound (and specifically, Machine Listening) be used for in a ‘dystopian’ scenario?

For each question, participants were asked to debate among themselves in order to reach some consensus on the theme, with their comments to be shared later with the entire group, and with non-expert participants during the open-door session. The conversation among each group was transcribed in real-time (using the online collaborative environment Miro) by a member of the Lab’s organising team who also served as a facilitator for the group. See Figure 2 for an example.

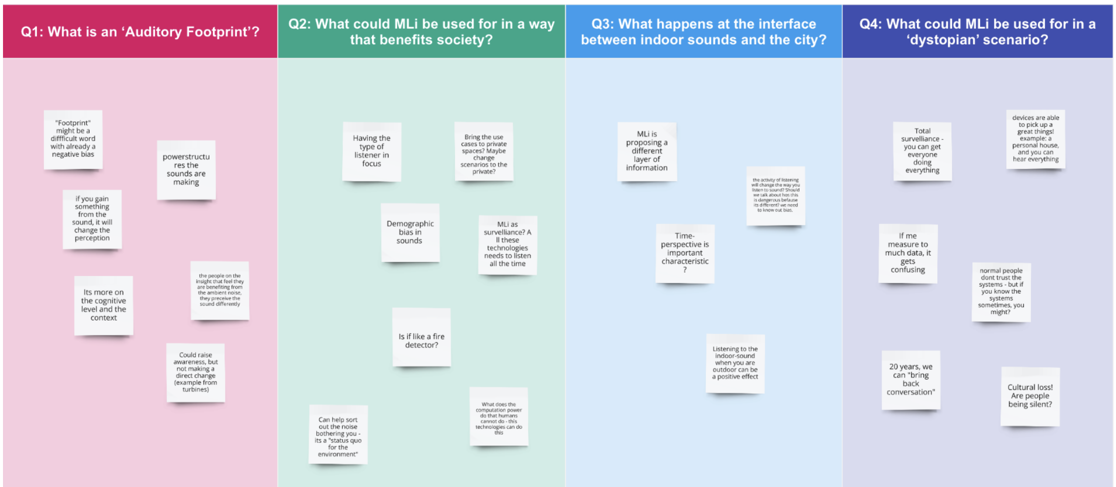

Figure 2. An excerpt of the Miro board representing the expert discussions.

2.1 Analysis of results

We report below on the discussion and insights provided by the participants for each trigger question.

Question 1: What is an “Auditory Footprint”?

In general, the experts from the three teams shared a similar viewpoint in highlighting that in the case of soundscape and sound events, ‘Footprint’ might be a difficult word to use. As citizens become more familiar with the concept of Carbon Footprint, the term might come "with already a negative bias" which, in turn, would fall short at conveying the cultural meaning that sound has in our personal and collective lives. As one group pointed out, the "cultural footprint of sound also reflects a structure of power" which, however, is not measurable as an aggregated index of "acoustic and physical properties of sounds events" - such as the Carbon Footprint is. Experts pointed at another potential ambiguity of the term since the word 'Auditory' refers directly to human perception and our cognitive abilities - again, unlike Carbon Footprint, the perceptual qualities of sound are not directly measurable as an objective property. Moreover, as humans, we are used to "filter out the noise" and focus on sounds that are meaningful to us. This brings an additional difficulty to the measurement of Auditory Footprints to identify the harmful component of the environment: in fact, the negative impact of a specific sound is highly contextual and - to some extent - subjective. As one group pointed out, "if you gain something from a sound (e.g., loud and upbeat music during a concert that empowers you), it will change your perception" and "if people feel they are benefiting from the ambient noise, they perceive the sound differently".

The discussion on the term Auditory Footprint highlights an underlying tension between the negative and positive impact that sound can have on listeners. Sound quality perception is contextual and to some extent subjective (versus the objective negative connotation of the Carbon Footprint) thus influencing how much we would act to minimize or change the impact of its 'footprint'. Additionally, while, once we are aware of our individual Carbon Footprint, there are measures we can take (or even buy) to balance it; an increased awareness of the negative impact of a specific sound source (e.g., noise coming from wind turbines) would not be easy to convert into direct action. The collective dimension of the soundscape and our individual agency can be in stark contrast. Sound has an inter-relational dimension where the causes of sound and its consequences are entangled - and individual responsibilities are difficult to delineate.

Question 2: What could MLi be used for in a way that benefits society?

The experts proposed several ideas on how MLi applications could be used to serve communities, with a focus on individuals affected by impairments. Yet, several experts pointed out the risks of the infrastructure required for MLi to be leveraged for surveillance applications, which have often debatable benefits to societies. Current AI efforts (both for analysis and content-generation) are becoming more invasive in our lives and are changing the ways we interact with and trust the world. Emerging efforts with SoundAI should consider the lessons learnt and offer more responsible and ethical solutions that benefit the entire society and the environment by addressing the real needs of people, other species and natural ecosystems, i.e., planetary wellbeing in general. Pros and cons of some application areas are presented below.

Tracking specific sounds through time in specific contexts could be an efficient way to understand certain societal and environmental challenges. In this case, "*using MLi would enable humans to discover subtle variations in sound patterns*" inaudible to the human ear- thus measuring hidden footprints in complex soundscapes. A "*city with ears*" could "*listen to* *the communities that are served by these systems*" (e.g., via widespread networks of microphones or audio sensors). This could lead to the development of an augmented collective hearing which could be used by individuals and groups to compensate for the lack of other senses, provide accessibility for the hearing impaired e.g., through "*an augmented 'instant listening' device, a sort of Soundscope, which would help people to magnify certain sounds and filter out unwanted noise*". It could also help identify hidden patterns of other-than-human species within the urban space, such as animals and pets. The computation power of MLI could "*do what humans cannot do: move the point of listening between ambient and targeted listening, support 'remote listening' for specific environments, act as a 'fire detector' to alert on specific sound triggers, an create a shared space in-between public and private soundscapes"*. Through massive audio data collected by MLI we could simulate the behaviour of specific soundscapes and how these might lead to unwanted consequences (e.g., traffic accidents at specific urban sites).

However, such technology could easily "*backfire, and instead allow for surveillance - such as acoustical cameras used to identify gunshot in several cities in the U.S., which ended up discriminating against specific communities*". This might endorse a demographic bias connected to specific sound events or soundscapes. Additionally, MLi technologies will have to "*listen all the time"* - how this would impact personal and collective rights both in public and private spaces must be a matter of concern. Lastly, before advocating for the massive introduction of AI technologies, a reflection should be conducted on the question "*what other ways could be leveraged to use current technology without adding another new technology?"*.

Question 3: What happens at the interface between indoor sounds and the city?

First and foremost, the experts agreed that when it comes to our everyday soundscapes, indoor and outdoor may not have a fixed boundary. From one side, there is a certain variable porosity that has to do with intimacy, privacy, and the psychological dimensions of private versus public spaces. At the same time, on the other side, indoor and outdoor could be treated as the acoustic equivalent of "*water and oil'*" - two substances that coexist but never really merge into each other. In situations where an "*inside (i.e., a private, personal space of comfort) does not exist, (for example, in hospitals, or in temporary housing–such as while traveling or renting i.e., when we are not 'at home'–or in extreme homeless situations*"), being able to create one's own soundscape by means of generative SoundAI might be both an individual and a collective source of wellbeing. Through sound, we could "*restore health for people with negative psychological conditions*", we could create our individual or collective "*mixtapes, with birds singing at night if that comforts us*", or "*listen to indoor sounds when we are outdoors*", or to feel like "*we are not lonely in the building"*. On the other hand, by means of Machine Listening applications, the interface between indoor and outdoor - an intangible acoustic border - could have a filter that cancels some specific sound.

Regardless of being indoors or outdoors, an important topic arose in the discussions that indicates that privacy, data ownership, and psychological safety are major topics of concern for advancing AI applications for sound. Where are, then, the 'acoustic' borders for acceptable personal, private, communal, and social soundscapes? Who do the sound events belong to? When do we give consent, and when is our speech and other sounds made public?

Question 4: What could MLi be used for in a 'dystopian' scenario?

The experts criticised that as a technology, the impact of Machine Listening on the individual and the collectives is neutral - the focus should be on "*who is in charge*" of the design, implementation and deployment in the public space of MLi solutions. In many contexts (for instance, when vulnerable groups are involved) "*sound might be the gate for the Trojan horse to enter*", leading to discriminatory policies and unfair decision-making. In extreme cases, a "*total surveillance*" scenario can be easily imagined where MLi devices are able to "*pick up our activity at home*" and systematize control or record our conversations so that they can be "*brought back 20 years from now*". What will happen if individuals - or entire communities - "*are being silent?*" We could experience cultural loss: the disappearance of the meanings of sounds that are not interpreted by systematically deployed Machine Listening.

There is a broader issue when trying to imagine MLi employed at large scale: that is, the analogy we tend to use is of image- and visual-based AI. In fact, sound is a different "*raw material*" altogether: While image is "*containable*", sound is "multifaceted from inception" in that it presents multiple semantic layers. How the multi-layered nature of sound will be transformed into discrete units for AI to interpret and forecast its behaviour, will be critical to the impact of MLi (and MLi - based solutions) on individuals and society at large. Thus, the question brought in accountability for responsible and ethical design decisions at the core of SoundAI and its introduction and application to society.

3. Reflection on the Lab

In general, a short-circuit emerged with the definition of the term Auditory Footprint. While it was originally intended as a concept that was primarily focussed on the listener, and holding a neutral connotation, the analogy with the term 'Carbon Footprint' seemed to instead put the focus on the emitter, and in a negative sense. More generally, the meaning of "footprint" is ambiguous. Footprints can be intended as the following:

- Feedback of an activity can be expressed as a footprint. In that sense, an Auditory Footprint would focus on the act of listening to decode and interpret the activity that generated it.
- An Auditory Footprint could also be interpreted as a measure of impact. However, it is unclear how the direct impact of a sound (e.g., because of a

specific sound, all whales in a particular area died) could be measured (as we do with the Carbon Footprint).

- Listening is non-univocal, i.e., it has a subjective component as it is always filtered by the characteristics (physical, biological, cultural) of the individual listener. An Auditory Footprint should account for such diversity in listening, and we should be able to automatise this process. In a future perspective, Auditory Footprints may move from the notion of listeners to a 'multi-listener level'.
- It is still unclear whether an Auditory Footprint is an aggregation of different processes (e.g., decoding the acoustic components, the perception of a sound, the identification of the source…), in the same way that a Carbon Footprint stands for an aggregated value of methane and carbon dioxide emissions as they impact the atmosphere.

While the Auditory Footprint as a concept does necessarily include the listener's perception (i.e., receiver), a Carbon Footprints is only focussed on the producer i.e., the measure of the environmental impact of products, services, and so on. Accordingly, the authors wondered whether '*Auditory Fingerprints'* could be a better metaphor for the intended function of Auditory Footprints. Fingerprints are unique for each person, as are the traces of sounds on the environment and on the listener. While it would require more custom care to develop MLi algorithms that sensitively consider the individual experience of sound and how it is influenced by contextual factors, identifying and classifying the *fingerprint* of a specific sound event could be a *good enough* descriptor to assess the impact of specific sound sources on a specific acoustic environment. The difference between Auditory Footprints and Fingerprints reflects the experts' critical approaches to the acts of listening that emerged during the Lab. When considering the potential subjectivity of Auditory Footprints, the experts highlighted that we could listen to the same sound various times (or in different contexts) and attribute different meanings to it. This could be exploited as a resource in designing automated listening toward MLi algorithms that are able to differentiate between listening and merely detecting sound.

Amongst other emerging themes, the most common – which deserve further attention in the next phases of our research – are:

1. *Listening to other-than-humans* (animals, pets, the natural world). Remote listening to natural phenomena could be a positive advancement in many fields, including monitoring climate change or biodiversity loss. MLi shouldn't be thought of as limited to the audible frequencies of humans.
2. *More-than-human listening*. MLi should allow for augmented listening or bionic hearing, and potentially also be able to abstract sound from its time-based nature. In the future, MLi applications could allow for listening to personalised 'microscapes' instead of soundscapes.

3. *Training dataset for MLi* is still a critical issue. In developing them, we should be wary of how and by whom audio datasets are created. An ethical issue could be how the perception of the perceived quality of the sound e.g., of its Auditory Footprint, would change if people were paid to listen to certain sounds for labelling purposes.
4. *MLi holds potential to increase the accessibility of sound* by fostering augmented hearing for the hearing impaired and by leveraging the restorative and comforting power of sound in cases of detrimental psychological conditions and unwanted loneliness.
5. *Finally, ethical implications emerged as having critical importance*. Several questions were raised that deserve urgent attention as AI applications of sound are on the increase as we write. MLis could become a way to target specific communities; and listening biases from the algorithm's designers could be embedded in the machines (e.g., a bias in soundscape studies that could be transferred to the algorithm). The definition of privacy is critical to prevent 'stigma by sound' and should be taken into account. Lastly, since people tend to have private conversations outdoors, the differences between collecting outdoor and indoor recordings as a mitigating measure (common in soundscapes studies) does not in reality solve privacy concerns.

4. Conclusions and Future work

In this Lab, we have discussed opportunities and threats that the application of AI to sound pose to society. Through brainstorming questions with nine experts and consolidations with a cohort of circa fifty design researchers we have developed five emerging themes that focus on *what and for whom* MLi can listen to, and the ethical implications that a widespread implementation of MLi technology in cities could lead to. As the application of MLi to research and design domains is still in its infancy, the Lab also focused on agreeing on a common terminology to characterize the different concepts related to MLi applications, some of which might be heavily subjective. Future work points in two directions: First, we aim to formalize the insights provided during the Lab into a framework and dictionary of definitions that could help structure research efforts in this domain. Second, we aim to discuss the feasibility of developing and implementing MLi applications in cities, looking both at technology and policy aspects.

It is our intention that SoundAI responds to the real needs of the individual humans, other species, and natural habitats in order to make positive societal and environmental impacts. However, there are unresolved issues regarding responsible SoundAI, including data ownership, accountability, and ethics. In order to avoid the privacy and data ownership concerns, focussing on other species could be an easy start to explore the needs of the algorithm and its implications on the MLi infrastructures (e.g., sensor technologies, storage, distribution). Lastly, in the Lab, not much attention was paid to the hard and soft technologies behind SoundAI. This will be the focus of future work as further investigation is needed to assess the feasibility of some of the solutions discussed by the Lab's participants.

Acknowledgements: The authors would like to thank all invited experts and participants of the DRS2024 Lab who joined us in person at Northeastern University in Boston and who dedicated their time and effort to make this Lab a success. This Lab was financially supported by the MIT-MISTI fund granted to the collaboration between MIT Senseable City Lab and Critical Alarms Lab (TU Delft). Dr Lenzi's work was partially funded by the Marie Skłodowska-Curie Cofund Programme of the European Commission (H2020-MSCA-COFUND-2020-101034228-WOLFRAM2).

5. References

Axelsson, Ö, Nilsson, M. E., and Berglund, B. (2010). A principal components model of soundscape perception. J. Acoust. Soc. Am. 128, 2836–2846. doi: 10.1121/1.3493436.

Bello, J. P., Silva, C., Nov, O., Dubois, R. L., Arora, A., Salamon, J., ... & Doraiswamy, H. (2019). Sonyc: A system for monitoring, analyzing, and mitigating urban noise pollution. Communications of the ACM, 62(2), 68-77.)

Brocolini, L., Lavandier, C., Marquis-Favre, C., Quoy, M., & Lavandier, M. (2012, April). Prediction and explanation of sound quality indicators by multiple linear regressions and artificial neural networks. In Acoustics 2012 (pp. 2121-2126), Nantes, France.

Calleri, C., Shtrepi, L., Armando, A., & Astolfi, A. (2018). Evaluation of the influence of building façade design on the acoustic characteristics and auditory perception of urban spaces. Building Acoustics, 25(1), 77-95.

Caniato, M., Bettarello, F. & Gasparella, A. (2021). Indoor and outdoor noise changes due to the COVID-19 lockdown and their effects on individuals' expectations and preferences. Sci Rep 11, 16533. https://doi.org/10.1038/s41598-021-96098-w

Chion, M. (1994). The Audiovision: Sound on Screen. New York, U.S.: Columbia University Press.

Duarte, CM, Chapuis, L, Collin, SP, Costa, DP, Devassy, RP, Eguiluz, VM, Erbe, C, Gordon, TAC, Halpern, BS, Harding, HR, Havlik, MN, Meekan, M, Merchant, ND, Miksis-Olds, JL, Parsons, M, Predragovic, M, Radford, AN, Radford, CA, Simpson, SD, Slabbekoorn,H, Staaterman, E, Van Opzeeland, IC, Winderen, J, Zhang, X and Juanes, F. (2021). The soundscape of the Anthropocene Ocean. Science: 371, eaba4658.

European Parliament. (2000). Directive 2000/14/EC of the European Parliament and of the Council of 8 May 2000 on the Approximation of the Laws of the Member States Relating to the Noise Emission in the Environment by Equipment for Use Outdoors. Available online at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32000L0014

Farina, A., Krause, B., Mullet. T.C. (2024). An exploration of ecoacoustics and its applications in conservation ecology, BioSystems, Volume 245. https://doi.org/10.1016/j.biosystems.2024.105296.

Gaver, W. W. (1993). What in the world do we hear? An ecological approach to auditory event perception. Ecological Psychology, 5(1), 1–29. https://doi.org/10.1207/s15326969eco0501_1

Javahid, S. (2023). Audio Data Collection for AI: Challenges and Best Practices in 2023. Available online at: https://research.aimultiple.com/audio-data-collection/ (accessed January 27, 2023).

ISO 12913-1. (2014). Acoustics—Soundscape—Part 1: Definition and Conceptual Framework. Geneva: International Organization for Standardization.

ISO/TS 12913-3. (2019). Acoustics — Soundscape — Part 3: Data Analysis. Geneva: International Organization for Standardization.

Lenzi, S., Ciuccarelli, P., and Offenhuber, D. (2024) Towards a Definition of Autographic Sonifications: Listening as an Act of Knowledge, in Gray, C., Ciliotta Chehade, E., Hekkert, P., Forlano, L., Ciuccarelli, P., Lloyd, P. (eds.), DRS2024: Boston, 23–28 June, Boston, USA. https://doi.org/10.21606/drs.2024.729

Lenzi, S., Spagnol, S., Özcan, E. (2023). Improving the quality of the acoustic environment in neonatal intensive care units: a review of scientific literature and technological solutions. Front. Comput. Sci., 05 October 2023 Sec. Human-Media Interaction Volume 5 – 2023. https://doi.org/10.3389/fcomp.2023.1156693

Lenzi S, Sádaba J, Lindborg P. Soundscape in Times of Change: Case Study of a City Neighbourhood During the COVID-19 Lockdown. Front Psychol. 2021 Mar 24;12:570741. doi: 10.3389/fpsyg.2021.570741.

Liu, X, Dong, J and Jeon, M. (2023). Robots' "Woohoo" and "Argh" can Enhance Users' Emotional and Social Perceptions: An Exploratory Study on Non-Lexical Vocalizations and Non-Linguistic Sounds. J. Hum.-Robot Interact. https://doi.org/10.1145/3626185.

Merino-Martinez, R, Pieren, R., Schäffer, B., & Simons, D. G. (2022). Psychoacoustic model for predicting wind turbine noise annoyance. In the proceedings of the 24th Congress in Acoustics. 24-28 October 2022. Gyeongju, Korea. https://www.dora.lib4ri.ch/empa/islandora/object/empa:31885

Mehta, R., Zhu, R., J., and Cheema, A. (2012). Is Noise Always Bad? Exploring the Effects of Ambient Noise on Creative Cognition. Journal of Consumer Research 39, no. 4 (2012): 784–99. https://doi.org/10.1086/665048.

Mitchell, A., Oberman, T., Aletta, F., Kachlicka, M., Lionello, M., Erfanian, M., et al. (2021). Investigating urban soundscapes of the COVID-19 lockdown: a predictive soundscape modeling approach. J. Acoust. Soc. Am. 150, 4474. doi: 10.1121/10.0008928

Mitchell, A. J. (2022). Predictive Modelling of Complex Urban Soundscapes: Enabling an engineering approach to soundscape design. Doctoral thesis (Ph.D), UCL (University College London).

Nautsch, A., Jiménez, A., Treiber, A., Kolberg, J., Jasserand, C., Kindt, E., et al. (2019). Preserving privacy in speaker and speech characterisation. Comput. Speech Lang. 58, 441–480. doi: 10.1016/j.csl.2019.06.001

Orhan, C. (2019). A comparative study on indoor soundscapes in museum environments (Master's thesis, Ekonomi ve Sosyal Bilimler Enstitüsü).

Özcan E., Broekmeulen C.L.H., Luck Z.A., van Velzen M., Stappers P.J., Edworthy J.R. (2022). Acoustic Biotopes, Listeners and Sound-Induced Action: A Case Study of Operating Rooms. International Journal of Environmental Research and Public Health. 19(24):16674. https://doi.org/10.3390/ijerph192416674.

Özcan, E., & van Egmond, R. (2012). Basic semantics of product sounds. International Journal of Design, 6(2), 41-54.

Özcan, E. (2014). The Harley Effect: Internal and external factors that facilitate positive experiences with product sounds. Journal of Sonic Studies. doi: 10.22501/JSS.242114

Phillips, JN, Termondt, SE and Clinton, FD. (2021). Long-term noise pollution affects seedling recruitment and community composition, with negative effects persisting after removal. In Proc. R. Soc. http://doi.org/10.1098/rspb.2020.2906.

Pedersen, K., Gee, K. L., Transtrum, M. K., and Butler, B. A. (2018). Machine learning-based prediction of outdoor ambient sound levels: ensemble averaging and feature reduction. J. Acoust. Soc. Am. 144, 1791–1791. doi: 10.1121/1.5067908

Salamon, J., Bello, J. P., Farnsworth, A., Robbins, M., Keen, S., et al. (2016). Towards the automatic classification of avian flight calls for bioacoustic monitoring. PLoS ONE 11:e0166866. doi: 10.1371/journal.pone.0166866.

Salamon, J., and Bello, J. P. (2017). Deep convolutional neural networks and data augmentation for environmental sound classification. IEEE Signal Process. Lett. 24, 279–283. doi: 10.1109/LSP.2017.2657381

Salazar Miranda, A., Halpern, D., and Duarte, F. (2020). Sonic Cities: Listening to Parks During Lockdown. Senseable City Lab, Medium. Available online at: https://tinyurl.com/26f5ernr. Last accessed: 21 Jan 2025.

Sevillano, X., Socoró, J. C., Alías, F., Bellucci, P., Peruzzi, L., Radaelli, S., et al. (2016). DYNAMAP–development of low cost sensors networks for real time noise mapping. Noise Mapp. 3, 172–189. doi: 10.1515/noise-2016-0013

Spagnol, S., Vila, N.V., Akdag Salah, A., Goos, T., Reiss, I., and Özcan, E. (2022). Towards a quieter neonatal intensive care unit: Current approaches and design opportunities, in Lockton, D., Lenzi, S., Hekkert, P., Oak, A., Sádaba, J., Lloyd, P. (eds.), DRS2022: Bilbao, 25 June - 3 July, Bilbao, Spain. https://doi.org/10.21606/drs.2022.258.

Spagnol, S., Goos, T. G, Reiss, I. and Özcan, E. (2022). An Algorithm for Automatic Acoustic Alarm Recognition in the Neonatal Intensive Care Unit. 7th International Conference on Frontiers of Signal Processing (ICFSP. pp. 59-63, doi: 10.1109/ICFSP55781.2022.9924684

Sueur, J. and Farina, A. (2015). Ecoacoustics: the Ecological Investigation and Interpretation of Environmental Sound. Biosemiotics 8, 493–502. https://doi.org/10.1007/s12304-015-9248-x

Torresin, S., Albatici, R., Aletta, F., Babich, F., Oberman, T., Stawinoga, A. E., & Kang, J. (2021). Indoor soundscapes at home during the COVID-19 lockdown in London–Part I: Associations between the perception of the acoustic environment, occupantś activity and well-being. Applied Acoustics, 183, 108305.

Torresin, S., Albatici, R., Aletta, F., Babich, F., & Kang, J. (2019). Assessment methods and factors determining positive indoor soundscapes in residential buildings: A systematic review. Sustainability, 11(19), 5290.

About the Authors:

Sara Lenzi is senior research fellow at Ikerbasque Basque Foundation for Science and Universidad de Deusto, Faculty of Engineering in Bilbao, Spain. Her main research interests cover data sonification, soundscape studies, and the definition of design tools, processes and methods in the context of human-data interaction. She conducted research on the changes in the urban soundscape during the COVID pandemic. She is the main applicant or co-applicant of several research projects on the automatic assessment of the quality of the soundscape in indoor environments. She has conducted more than 40 workshops and soundwalks intended to increase awareness of the impact of sound in everyday life with sighted and non-sighted participants.

Elif Özcan (she/her) is associate professor and director of Critical Alarms Lab at TU Delft, NL. Her expertise lies in sound-driven design by which sound provides issues and/or opportunities for any type of design ranging from sound design to acoustics engineering, from interaction design to data-centric system design. She is one of the founders of the DRS SIG Sound-driven Design. Her recent interests focus on understanding acoustic environments, soundscapes and acoustic biotopes to be able to unravel the relationship between sound-induced action and the listening experience.

Simone Mora is a research scientist at MIT Senseable City Lab and an adjunct associate professor at NTNU Norway. He does research on low-cost and opportunistic strategies and ideation methods and tools to support their development. He has experience organizing and facilitating workshops in multidisciplinary settings. He recently co-organized a workshop on hybrid co-design methods and he co-organized the 'Data Slots' lab at DRS 2022.

Martina Mazzarello is a postdoctoral researcher at MIT Senseable City Lab. With a PhD in Spatial Urban Design from Politecnico di Milano, Martina's research focuses on integrating physical, digital, and human interactions within urban environments. She is currently leading research projects using data and science to better understand changes in ways of living and working in cities. She co-organized the 'Data Slots' lab at DRS 2022.

Åse Håtveit is a visiting student at the MIT Senseable City Lab from the Norwegian University of Science and Technology, where she is conducting research for her master's thesis in informatics. Her current research focuses on the challenges associated with the development and adaptation of urban environmental sensing projects for makers, as well as on the potential for engaging communities through the use of open-source, low-cost sensing technologies to better understand their environments.

Fábio Duarte is a principal research scientist at MIT Senseable City Lab and lecturer in the MIT Department of Urban Studies and Planning. He co-authored *Urban Play: make-believe, technology and design* (MIT Press), and has published in journals including *Nature Cities, Nature Sustainability, and Scientific Reports*, among others. He is one of the co-authors of 'Data Slots'.