Sound and Safety in the Age of Electric Vehicles

Pedestrian Insights from the EVA Project

Flaithrí Neff
The IDEAM Research Institute
Technological University of the Shannon
Limerick, Ireland
e-mail: Flaithri.Neff@tus.ie

Linda Butler
School of Education, Hibernia College
Dublin, Ireland
lbutler@hiberniacollege.net

Dónal Fitzpatrick, Gerald Craddock, Carly Cheevers, Aislinn Conway
The Centre for Excellence in Universal Design
The National Disability Authority
Dublin, Ireland
donal.fitzpatrick@nda.ie
gerald.craddock@nda.ie
carly.cheevers@nda.ie
aislinn.conway@nda.ie

***Abstract* — Electric vehicles (EVs) pose challenges for pedestrian safety due to reduced sound levels at low speeds, prompting regulations mandating Acoustic Vehicle Alerting Systems (AVAS). This article reports findings from a survey of pedestrians with and without vision-impairment. Participants with vision-impairment reported significantly lower perceptions of safety, detectability, and interpretability, while both groups frequently described surprise encounters with EVs. These findings expose three persistent challenges in AVAS regulation: the absence of distance-based design guidance beyond basic sound pressure thresholds; a tendency toward branding-focused rather than functionally informative sound-design; and the uncertain effects of overlapping AVAS signals in future urban soundscapes. More broadly, these issues reflect issues around limited academic engagement in standards and regulatory development, where commercial feasibility has often outweighed perception-research evidence. We argue that future progress depends on embedding pedestrian-centered metrics into standards, allowing psychometric research to inform compliance testing in ways that more accurately reflect real-world pedestrian experience.**

Keywords- Electric Vehicles; Acoustic Vehicle Alerting Systems; Pedestrian Safety; Sound Perception.

I. INTRODUCTION

An electric vehicle (EV) is a mode of transport that is powered by electricity rather than conventional internal combustion fueled by petrol or diesel. It employs an electric motor driven by a rechargeable battery, which can be replenished using dedicated charging stations or an installed home/business electrical outlet. EVs are typically categorized into three main types: 1) Battery Electric Vehicles (BEVs), which are fully electric and contain no internal combustion engine (ICE) components; 2) Plug-in Hybrid Electric Vehicles (PHEVs), which combine an electric motor with a petrol engine, allowing limited electric-only operation before switching to fossil fuel; and 3) Hybrid Electric Vehicles (HEVs), which also include both an electric motor and an ICE, but differ from PHEVs in that the battery is not recharged via an external source but rather through regenerative braking and engine operation.

The primary advantages of EVs include significantly lower tailpipe emissions (in some cases zero), reduced operating costs, lower maintenance requirements due to fewer moving mechanical components, and very quiet operation. It is this significantly-reduced noise output, particularly at low speeds, that has introduced new safety concerns [1][2][3][4] - especially for vision-impaired pedestrians who rely on auditory cues to detect oncoming traffic due to having reduced vision or complete sight-loss. In response, electric and hybrid vehicles are now equipped with an Acoustic Vehicle Alerting System (AVAS), which emits artificial sound to announce the vehicle's presence to nearby pedestrians and cyclists when operating at low speeds (see Section II for information on why AVAS initiates only at low speeds).

Nonetheless, empirical studies consistently show that pedestrians exhibit delayed reaction times and inaccurate distance judgments when engaging with EVs compared to ICE vehicles [5][6][7][8][9][10][11][12]. While AVAS-generated sounds have been shown to improve pedestrian detection rates compared to earlier near-silent EV models, evidence also suggests that current AVAS signatures lack certain acoustic cue-information that is naturally present in conventional ICE vehicle sounds [13]. Therefore, as EV adoption continues to accelerate [14][15], there is a growing imperative for further research into the acoustic characteristics of both ICE acoustics and AVAS, including how they influence pedestrian perception, interaction, and overall safety-outcomes in diverse real-world contexts.

While regulatory frameworks define broad acoustic characteristics and compliance thresholds for AVAS-equipped vehicles, there remains a significant knowledge gap regarding how effective AVAS performs at supporting real-world vehicle localization. Indeed, AVAS standards treat distance as a geometric parameter, but they do not determine whether such distances correspond to pedestrians' perception of a safe margin. In making a small contribution to closing such gaps, we present the EVA (Electric Vehicle Acoustics) study where we set out to gather experiential data from two key pedestrian cohorts: individuals without vision-impairment and those with vision-impairment. Using this comparative approach, our goal is to show how AVAS signals are perceived in practice by pedestrians who can rely on visual cues to compensate for potential AVAS limitations versus those who cannot. This information, therefore, helps evaluate whether current standards and regulations sufficiently account for the needs and experiences of all pedestrians - particularly those with vision-impairments.

In presenting these findings and their implications for standards and policy, the article is organized as follows. Section II introduces the regulatory frameworks that underpin AVAS and Section III explains the methodology adopted for implementing the EVA survey and analyzing its data. Sections IV to VI present the quantitative findings step by step: first a descriptive analysis of the Likert items, then their correlations, and finally the results of significance testing. Section VII turns to the qualitative material, analyzing participants' open-ended responses. Section VIII brings these strands together in discussion, connecting the EVA findings to wider regulatory debates and highlighting the persistent perceptual gaps in AVAS. Section IX sets out the main conclusions, and Section X looks ahead to future work, particularly the development of perception-based metrics, with a final statement on how future AVAS may progress as enabling-technologies come on stream – producing for example context-aware, AI-enhanced AVAS systems.

II. AVAS Regulatory Frameworks

In the European Union, the inclusion of AVAS became a legal requirement for all newly approved electric and hybrid vehicles as of 1 July 2019, under Regulation (EU) 2019/2144 [16]. This regulation amends the earlier Regulation (EU) No. 540/2014 [17], which governs permissible noise levels from motor vehicles. The technical parameters for AVAS design and performance are based on the United Nations Economic Commission for Europe (UNECE) Regulation No. 138 [18], which has been transposed into EU law through the above regulatory instruments. Of note is that from 1 July 2021, the AVAS mandate was extended to all newly sold electric and hybrid vehicles within the EU, which was a transitional period intended to allow the vehicle industry to adapt to the new requirements. Since these regulations are not retroactive, many EVs manufactured before 2019 remain road-legal even though they lack any AVAS capability and are effectively silent at low speeds in built-up urban areas. In practice, this means that even vehicles produced as late as the first half of 2021 may still be on the road without AVAS installed, yet remain fully compliant with the law.

For those EVs that do come under these EU regulations, AVAS must activate automatically when the vehicle is moving at speeds below 20 km/h or when reversing. At speeds above 20 km/h, it is considered that the ambient noise produced by EVs is comparable to that of ICE vehicles, whereby the vast bulk of vehicular noise is dominated by tyre-on-road sound.

AVAS requirements currently apply to both vehicle categories M and N (passenger and goods transport vehicles), although increasing attention is also being directed toward smaller personal mobility devices (e.g., electric scooters), which pose similar risks but are typically regulated at municipal or national levels and lack a harmonized EU-wide framework.

To understand how AVAS requirements crystallized, it is useful to consider the minutes of UNECE's working groups - particularly the *Quiet Road Transport Vehicles* (QRTV) informal group (2010–2012) [19]. This group gathered inputs from national authorities, academics, industry, disability organizations, and trade representatives, balancing these perspectives and distilling a substantial body of third-party contributions. Responsibility then passed to the UNECE's *World Forum for Harmonization of Vehicle Regulations* (WP.29) [20], which oversaw the eventual publication of Regulation No. 138. Prominent inputs included surveys and market data from the Japan Automobile Standards Internationalization Center (JASIC); controlled trials by Germany's Bundesanstalt für Straßen- und Verkehrswesen (BASt); early U.S. research updates from the National Highway Traffic Safety Administration (NHTSA); and consolidated manufacturer positions via the Organisation Internationale des Constructeurs d'Automobiles (OICA). Importantly, these early discussions repeatedly highlighted perceptual concerns - such as detectability, locatability, attention-catching "departure" cues, and directivity - as well as scope debates (e.g., whether very quiet ICE vehicles or powered two-wheelers/e-bikes should be included).

As submissions progressed, however, some ambiguities also emerged in the way data were presented to the UNECE process. For instance, certain comparative trials reported EV detectability as equivalent to that of ICE vehicles, yet did not specify which ICE models were used as benchmarks. The absence of such detail makes it difficult to evaluate whether the chosen comparators may have been uncharacteristically quiet - as is the case with some luxury petrol vehicles - thereby potentially biasing the impression that EVs already achieved acceptable detectability levels as early as 2011. Such limited transparency may have reinforced a regulatory focus on minimal SPL thresholds, rather than encouraging a more comprehensive exploration of perceptual outcomes across varied real-world contexts. As a result, many of the recurring concerns raised during these formative discussions were only partially reflected in the final text of UNECE's 2016 regulation, which ultimately centered on minimum A-weighted sound levels, one-third-octave coverage, and a speed-correlated frequency shift - while leaving critical perceptual dimensions such as locatability and effective distance-estimation largely un-operationalized. Another illustration of certain compromises made during the regulatory process was the exclusion of Powered Two-Wheelers (PTWs), such as motorcycles, scooters, and micromobility devices from scope as early as 2011. This decision reflected what could be considered a commercially pragmatic narrowing of the regulatory frame-of-reference, prioritizing feasibility and industry consensus over the integration of dedicated, first-hand pedestrian studies in relation to PTWs.

Therefore, the UNECE - and by extension the European Union - did not build its regulation around a series of dedicated perceptual studies as did the NHTSA when devising U.S. AVAS regulation in line with the Pedestrian Safety Enhancement Act (PSEA) [21]. Nonetheless, the two regulatory pathways ultimately converged on broadly similar requirements, reflecting a shared emphasis on feasibility, measurability within existing acoustic test procedures, and commercial practicalities. While this alignment has ensured regulatory clarity and harmonization across jurisdictions, it has also meant that the richer potential of perceptual-based academic research has yet to be systematically integrated into either framework. Internationally, AVAS regulation has relatively few but sometimes notable differences [22]. In the United States, it is regulated by FMVSS No. 141 [23] and in China by GB/T 37153-2018 [24].

From a compliance-testing perspective, the methodologies used to assess AVAS performance are set out in ISO 16254 [25], a standard developed by the International Organization for Standardization (ISO). They provide a measurement framework rather than a pass/fail benchmark, as regulatory compliance is determined by whether a vehicle meets the thresholds specified in UNECE R138 and corresponding EU regulations, not by ISO 16254 itself. Instead, the ISO standard serves as the procedural basis for applying those criteria, ensuring harmonization across key acoustic properties such as sound pressure level (SPL), frequency content and tonality, as well as in the definition of pass-by measurement setups, instrumentation, and track/environmental specifications. Table I summarizes how ISO and UNECE complement one another in defining and enforcing AVAS compliance.

A. AVAS Specifications

The evolution of AVAS research reflects a trajectory from initial recognition of safety risks posed by silent vehicles [26][27][28] to the development of sophisticated sound-design models and compliance frameworks [29][30][31][32]. Indeed, this field is of considerable societal importance. As discussed in Section II, the rise in pedestrian accidents involving quiet vehicles has led to global legislation defining sound levels, spectral content, and activation rules, with the aim of improving detectability while limiting noise pollution [30][33][34]. Therefore, while manufacturers have flexibility to design their own sound signatures - which they often leverage for brand identity purposes - certain high-level criteria must be reached in accordance with the regulations. Table II outlines the key components of the AVAS regulation within the EU area.

B. Branding versus Function

Despite the presence of regulatory specifications, the real-world performance of AVAS continues to face growing scrutiny [10] [35][36]. Implementations differ markedly across manufacturers, and many AVAS designs fulfil a dual role: (1) providing a pedestrian safety warning and (2) contributing to the vehicle's sonic branding, whereby the sound symbolically conveys the identity or characteristics of the vehicle or manufacturer [37][38][39]. While existing regulatory frameworks seek to enhance detectability and limit annoyance [40][41][42], these goals may be compromised when AVAS signals are also shaped by branding requirements. In many cases, the resulting design choices are influenced more by aesthetic, marketing, or brand-identity considerations than by functional acoustic safety performance [43][44][45]. While aesthetic distinctiveness is recognized as a legitimate commercial objective, including within the regulatory framework, it cannot be allowed to undermine the functional performance required in safety-critical applications.

AVAS, therefore, offers an opportunity to strengthen interdisciplinary collaboration between creative professionals (e.g., composers, sound designers) and scientific experts in psychoacoustics, vehicle acoustics, and urban noise [34][46][47][48][49]. Such collaboration could ensure that aesthetic motivations do not inadvertently undermine AVAS' functional effectiveness in real-world pedestrian environments. Further examination may be needed to determine whether current regulatory frameworks and standardization processes adequately support such interdisciplinary engagement. In particular, it would be valuable to explore how recognizing the role of aesthetic branding could encourage greater inclusion of end-user perspectives, especially those of vision-impaired pedestrians.

As such, more inclusive, user-centered studies would provide critical empirical data to evaluate whether AVAS systems balance branding ambitions with perceptual clarity and functional reliability [50][51][52][53][54]. Indeed, although all AVAS signals in EU-category M and N vehicles must conform to specified thresholds and satisfy ISO-based compliance tests, the substantial design latitude afforded to sound designers has raised increasing concern over whether AVAS systems adequately meet the needs of pedestrians who rely on auditory cues [50][55]. Relatively few studies have engaged directly with vision-impaired pedestrians, leaving regulators and manufacturers with only a limited understanding of how these groups perceive and interpret different AVAS designs.

While UNECE R138 and ISO 16254 embed some psychoacoustic concepts (particularly tonality and modulation) into regulatory and testing frameworks [56]; they do so without explicitly and directly referencing some of the newer scientific literature that underpins those conventions. Therefore, it is perhaps legitimate to question how such regulatory requirements have been codified and subsequently operationalized, particularly when manufacturers face competing pressures between branding differentiation and compliance. Essentially, despite regulatory progress, a critical knowledge gap remains regarding the optimal balance between regulatory compliance, psychoacoustic effectiveness, and inclusive end-user acceptability [57][58][59][60].

TABLE I. ISO 16254:2024 MEASUREMENT PROCEDURES AND UNECE REGULATION NO. 138 AVAS REQUIREMENTS. COMPARED TO ITS INITIAL 2016 EDITION, ISO 16254:2024 EXPANDS THE ACOUSTIC TESTING FRAMEWORK BY INCORPORATING ADDITIONAL MICROPHONES FOR PASS-BY DATA CAPTURE

Dimension	ISO 16254:2024 (Measurement Standard)	UNECE Regulation No. 138 (Regulatory Thresholds)
Purpose	Defines standardized methods for measuring AVAS sounds in controlled environments.	Sets mandatory performance thresholds and operational conditions for AVAS systems in electric and hybrid vehicles.
Scope	Applicable to M and N category vehicles at standstill and forward motion up to 20 km/h (also includes reverse).	Applies to M and N category vehicles; AVAS must operate up to 20 km/h and in reverse gear.
Minimum Sound Pressure Level	Measures A-weighted SPL[a] using 10-mic array (or full hemisphere in some setups); includes FFT[b] and psychoacoustic metrics.	Must emit at least 50 dB(A) at 0 km/h (stationary); at least 56 dB(A) at 10 km/h; at least 61 dB(A) at 20 km/h
Frequency Content	1/3-octave band spectrum[c] from 160 Hz to 5000 Hz is measured and reported	Must contain at least two 1/3-octave bands between 315 Hz and 5000 Hz, one of which must have the maximum SPL
Frequency Shift Requirement	Tonal shift (Δf) measured across speed; FFT used to evaluate frequency increase from 5 to 20 km/h	AVAS must show a frequency shift correlated with vehicle speed between 0 and 20 km/h; this can be achieved by modulation or increasing pitch
Tonality	Calculates tonal loudness, modulation, and tonality metrics (e.g., Zwicker method [56])	No numerical psychoacoustic threshold; requirement is that the sound must be recognizable and interpretable as vehicle motion by pedestrians (these are vague, whereas the updated ISO standard is now specific in what and how to measure)
Reverse Operation	Measures sound while reversing under standardized conditions	AVAS must emit sound when reverse gear is engaged, regardless of speed
Pause Function	Optional mute function noted in test description	Must not be continuously "pausable" by the driver; only temporary (e.g., for parking assist) interruptions are permitted
Microphone Configuration	10 microphones in linear or circular array (2 metres[d] from test line, 0.8–1.6 m height); also near-field mic for module testing, In the 2016 version of the standard, this comprised 2 microphones.	Test setup follows ISO 16254 (2016 version) - microphone position and track environment are indirectly inherited
Testing Environment	Outdoor (ISO 10844 compliant test track [61]) or indoor anechoic chamber; environmental conditions specified	Requires ISO 10844 test track or equivalent; temperature and wind speed constraints apply
Reporting Output	Full SPL and FFT data; metrics like $Lcrs_{10}$, Lst_{fwd}, Lst_{rev},[e] 1/3-octave spectrum plots	Pass/fail based on SPL thresholds, tonal content, and frequency shift behavior

a. Sound Pressure Level (SPL) is a ratio measure of the pressure variation in air caused by sound waves, relative to atmospheric pressure. When detected by a listener, the degree of this variation is perceived as loudness. Because SPL can be applied both in physical measurement and perceptual assessment, weighted scales are used to clarify what dimension is being addressed. When SPL is considered in relation to the minimum sound pressure required for human hearing, the A-weighted scale is typically used.

b. The Fast Fourier Transform (FFT) is a computational process that decomposes a captured and digitized acoustic signal into its constituent frequencies along the frequency domain. This decomposition reveals the individual frequency components that together form the complex waveform.

c. A 1/3-octave band spectrum is a way of grouping FFT information into frequency bands that are each one-third of an octave wide. An octave represents a doubling of frequency (for example, 500 Hz to 1,000 Hz) and dividing that range into three parts gives relatively narrow bands still useful for analysis.

d. The 2-meter measurement distance is a technical standard for repeatable testing, not a proxy for perceptual detection distance in real-world pedestrian contexts.

e. Shorthand for the measurement of SPL under certain vehicle conditions, where L=SPL; crs10 = constant running speed (10 km/h); stfwd = starting forward; strev = starting reverse.

TABLE II. IMPLEMENTATION OF UNECE INTO EU AVAS REGULATION

Action	Description
Automatic Activation	Must activate without driver intervention when the vehicle moves forward at speeds up to 20 km/h or when the vehicle is in reverse, regardless of speed.
Minimum Sound Pressure Levels (A-weighted SPL at 2 m)	50 dB(A) at 0 km/h (stationary, idling with AVAS on); 56 dB(A) at 10 km/h; and 61 dB(A) at 20 km/h
Frequency Content Requirements	Must include at least two 1/3-octave bands between 315 Hz and 5 kHz. One band must contain the maximum SPL while the second must be within 5 dB(A) of the maximum.
Speed-Dependent Frequency Shift	The signal must exhibit a perceptible increase in frequency (pitch or modulation) as the vehicle speed increases from 0 to 20 km/h. This shift must be continuous and clearly correlated with speed.
Reverse Operation Sound	When reversing, the AVAS signal must also operate and may be distinct from the forward sound. It must still meet minimum SPL and content requirements.
Tonal Quality and Recognisability	The sound must be recognizable as a vehicle in motion and it must be clearly audible in an urban background noise environment. No exact tonal or psychoacoustic metric is required in the current regulation, but tonality and modulation are implicitly expected to enhance detectability.
No Permanent Deactivation	A pause/mute function is not allowed unless it is temporary and is still required for specific operational scenarios (e.g., parking assist). The AVAS must resume automatically when the condition ends.
Compliance with Testing Protocols	Must be verifiable on an ISO 10844:2014 test track [61]. The new AVAS standard specifies additional requirements for outdoor test conditions, where no reflective surfaces can be within a 50m radius of the source and microphones. Measurement is performed with 10 microphones, placed at 2m distance and 0.8m-1.2 m height.

III. Methodology

The EVA study employed a mixed-methods survey to investigate pedestrian experiences and perceptions of EVs and AVAS, with particular attention to individuals who rely on auditory cues for navigation. The survey was ethically approved by the Ethics Committee of the Technological University of the Shannon prior to distribution.

Participants were recruited internationally by reaching out to road safety organizations, visual-disability advocacy groups, and academic forums in auditory science. Dissemination channels also included targeted mailing lists, disability support networks, and social media platforms. Eligibility criteria required participants to be 18 years or older, capable of providing informed consent, and to have previously encountered one or more EVs (either visually or aurally) in real-world environments.

Accessibility was a core design principle of the survey instrument. The online survey platform (SurveyMonkey®) was tested and optimized for compatibility with screen readers and other assistive technologies, ensuring independent completion by individuals with vision-impairments. Vision-impairment status was self-reported by participants in line with National Disability Authority (NDA) definitions [62], encompassing both full and partial vision loss. In addition, to broaden accessibility and comprehension, the survey language was structured in accordance with guidelines from the National Adult Literacy Agency (NALA) [63] to ensure individuals with varying literacy levels across both groups could understand and engage with the questions effectively. To safeguard privacy, no personally-identifiable information was requested, and IP tracking was disabled. Informed consent was embedded within the survey introduction, and final submission was treated as confirmation of consent.

A total of 135 individuals responded to the survey, with initial data cleaning removing 25 respondents who disengaged after initial consent, leaving a total of 110 legitimate responses. Participants were asked to self-report sensory impairments, if any - see Table III. With Hearing Sensory Impairment (HIS) and Other Stated Impairment (OSI) groups having insufficient sample sizes for comparative statistical analysis, these respondents were subsequently excluded. The final cleaned dataset, therefore, comprised 90 individual responses: 51 participants self-reporting no sensory impairment (NSI) and 39 self-reporting a vision sensory impairment (VSI). Data were stored in CSV format and statistically analyzed using R Version 4.4.2.

TABLE III. Respondent count broken into sensory categories.

Category	Number
No Sensory Impairment (NSI)	51
Visual Sensory Impairment (VSI)	39
Hearing Sensory Impairment (HSI)	17
Other Stated Impairment (OSI)	3
TOTAL	**110**

A. NSI and VSI Groups

Figure 1 shows the age distribution of participants in NSI and VSI groups. While the VSI group had a slightly higher median age compared to the NSI group, the ranges overlap substantially. This therefore indicated that age was unlikely to be a confounding factor in interpreting the differences in experiences or perceptions between groups.

The survey itself comprised two components:

- A series of Likert-scale statements assessing perceptions of EV detectability and AVAS effectiveness, which progressed from an analysis of a smaller subset of early respondents published in Neff et al. [1].
- Open-ended questions inviting qualitative insights into pedestrian experiences, which also progressed from an analysis of a smaller subset of early respondents published in Neff et al. [64].

This necessitated a mixed-methods analysis combining descriptive statistics and comparative analysis, along with thematic coding and sentiment analysis to identify key perceptual differences between the cohorts. Note that response patterns varied across sections (see Table IV), with 48 NSI and 39 VSI participants completing the Likert items, and 39 respondents from each group providing qualitative responses.

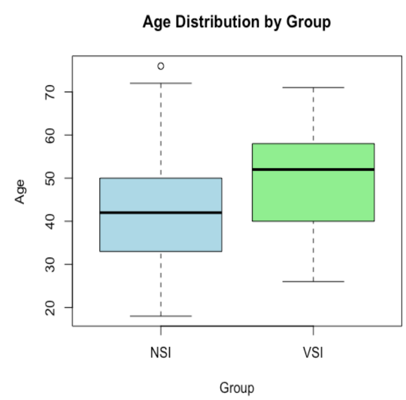

Figure 1. Age distribution of participants with NSI and VSI, shown as a boxplot. The VSI group shows a higher median age and narrower lower age range compared to the NSI group, which includes younger participants and one older outlier. The NSI Group's age range was 18 to 76 compared to the VSI Group's range of 26 to 71. The NSI Group's' age median was 42 compared to the VSI Group's median of 52; and the NSI Group's age mean was 42.44 compared to the VSI Group's mean being 49.06

TABLE IV. Number of participants for each survey section by group.

Group	Total Participants	Likert Responses	Open-Ended Responses
NSI	51	48	39
VSI	39	39	39

B. Likert Statements

To capture quantitative perceptions of EV and AVAS experiences, the survey included a set of nine structured statements (see Table V) assessed using a 5-point Likert scale.

These statements were designed to evaluate pedestrian views on EVs that they encountered and the detectability, interpretability, and overall effectiveness of AVAS. Participants were asked to indicate their level of agreement with each statement, reflecting their subjective experience and comfort in scenarios involving EVs.

Statements L1 through L8 used a common response scale as follows: 1 = I disagree a lot; 2 = I disagree just a little; 3 = I don't know; 4 = I agree just a little; 5 = I agree a lot.

Statement L9 maintained the same ordinal structure but was reworded to reflect perceived difficulty as follows: 1 = Very difficult; 2 = Difficult; 3 = Neither difficult nor easy; 4 = Easy; 5 = Very easy.

This approach allowed for the quantification of perceptual and behavioral dimensions of AVAS from both sighted and vision-impaired participants, enabling cross-group comparison. The inclusion of a neutral midpoint ("I don't know" or "Neither difficult nor easy") permitted non-binary responses and accommodated uncertainty where relevant.

TABLE V. THE NINE LIKERT STATEMENTS IN THE EVA SURVEY.

Code	Statement
L1	I feel safe when I think there might be an EV close by.
L2	It is easy to notice an EV approaching because of its sound.
L3	Sounds made by EVs help me understand what the vehicle is doing.
L4	I feel confident I understand an EV's next action based on its sound.
L5	I can react quickly to the sound of an EV when necessary.
L6	I find the sound of EVs pleasant.
L7	It takes little effort for me to listen to an EV's sound and understand what it is doing.
L8	I believe that the sound from all electric cars will be a positive thing for noise levels in busy cities and towns.
L9	Imagine you are standing on a busy street with lots of electric cars making sounds. Do you think it would be easy or hard to know when it is safe to cross the road?

C. Descriptive Analysis for Likert-Scale Items

Before conducting inferential statistical analysis, an initial descriptive analysis was performed to examine central response tendencies across the nine Likert-scale items. This was done to help determine if follow-on analysis was justified and also to contextualize overall response patterns that might highlight early group-level perceptual differences.

Given the ordinal nature of Likert data, the median serves as a more appropriate measure of central tendency than the mean. Therefore, each Likert item was calculated separately for the NSI and VSI groups, and responses were numerically coded on a 5-point ordinal scale, with consistent directional alignment across all statements (1 = most negative, 5 = most positive).

The median calculations were followed by a measure of the interquartile range (IQR), which provides an indication of the variability in participant responses within each group. Since normal distribution cannot be assumed in ordinal Likert responses, the IQR was deemed an appropriate non-parametric measure of statistical dispersion at the local group level. This served the purpose of giving the authors a vantagepoint of where the central 50% of responses fall (i.e., between the 25th and 75th percentiles), offering an indication of how consistent or varied participant ratings were for each statement. When interpreted alongside the median, the IQR allowed for a richer understanding of both central tendency and response spread.

In addition, effect size was calculated for each Likert item to estimate the magnitude of perceptual differences (if present) between the NSI and VSI groups. This was also done to estimated how substantial any inferential difference was in subsequent significance testing independent of sample size. Cliff's delta (δ) was selected as an appropriate non-parametric effect size measure, as again, it is suited to ordinal data. This statistical method quantifies the likelihood that a randomly selected participant from one group will have a higher or lower score than a participant from the other group. Values range from –1 to +1, where 0 indicates complete overlap, and values approaching –1 or +1 indicate stronger group-level divergence. When considered alongside medians and IQRs, Cliff's delta contributed to a more comprehensive understanding of group-level differences and highlighted which Likert items should undergo deeper inferential analysis.

D. Inferential Analysis for Likert-Scale Items

To move beyond descriptive summaries and examine whether observed group-level patterns reflect statistically meaningful differences, a series of inferential statistical analyses were conducted. These analyses aimed to explore both within-group associations and between-group differences in responses to the Likert items. This included both correlation analysis, significance testing, and multivariate analysis.

Correlation Analysis: To explore potential relationships between individual Likert-scale items within each participant group, Spearman's rank correlation coefficient was used. This non-parametric method assesses the strength and direction of monotonic associations between variables without assuming normal distribution. Separate correlation matrices were calculated for the NSI and VSI groups, allowing for the identification of within-group response patterns that may reflect underlying perceptual characteristics.

Significance Testing: To evaluate whether responses to Likert statements differed significantly between the NSI and VSI groups, Mann-Whitney U tests were performed for each Likert item. The Mann-Whitney U is a non-parametric test and suitable for independent samples and ordinal data, making it appropriate for comparing central tendencies between the NSI and VSI groups. In addition to p-values from the U tests, Rank-Biserial Correlation (r) was computed as an accompanying measure of effect size, offering insight into the magnitude and practical relevance of observed differences.

Multivariate Group Difference Analysis: To complement item-level testing and to examine overall response patterns across the full set of Likert items, a Permutational Multivariate Analysis of Variance (PERMANOVA) was conducted. This technique allowed us to compare group-level profiles across the multiple dimensions using non-parametric permutation tests. A Gower distance matrix was used to account for the ordinal nature of our Likert responses. As only two groups (NSI and VSI) were compared, the test included just a single degree of freedom in the analysis.

Taken together, these inferential analyses were selected to provide a robust approach to analyzing our Likert data whilst being careful about assuming normal distribution. All in all, they helped us to understand both the structure and statistical significance embedded in the differences or similarities between the NSI and VSI groups when it came to documenting their experiences and perception of AVAS.

E. Thematic and Sentiment Analysis of Questions 1 and 2

In addition to the Likert-scale items, the survey included four open-ended prompts designed to elicit qualitative insights, including the way these responses were conveyed on a sentiment level. These qualitative questions were specifically developed to capture perceptual nuances, emotional responses, and real-world encounters that Likert-scale items often fail to reveal clearly. These questions and their associated research rationale are presented in Table VI.

In R, the *tidyverse* suite was used for data preparation, and *rstatix* supported nonparametric testing. Thematic coding was conducted using the *quanteda* package, while *sentimentr* was used to calculate polarity scores and examine emotional tone. Therefore, analysis of the open-ended questions was multi-phased, relying on a hybrid computational-linguistic workflow focusing initially on only Questions 1 and 2. Question 1 was treated as a potential source of thematically rich, content-driven information about participants' baseline mental models of EVs versus ICE vehicles, whilst Question 2 served to provide insight into the affective tone characteristics of EV acoustics.

The first phase of analysis involved dictionary-based thematic coding of Q1 using predefined keyword sets aligned to six categories – namely: Comfort; Audibility; Environment; Cost; Performance; and Safety. These themes were chosen to map both functional concerns (e.g., detectability or environmental benefit) and experiential attitudes (e.g., acoustic pleasantness or risk). Each response was parsed for the presence of these theme indicators, resulting in binary (TRUE/FALSE) flags for each participant across all six dimensions. This allowed direct comparison of theme frequency between NSI and VSI groups.

The thematic categories were derived through an iterative process informed by two sources: (1) a review of prior AVAS and transport acoustics literature, which frequently organizes perceptual responses along functional and experiential dimensions, and (2) scanning of a subset of responses to identify recurrent terms and concerns. This dual approach ensured that the coding framework captured both established research themes (e.g., detectability, environmental benefit) and emergent participant attitudes (e.g., acoustic pleasantness, perceived risk).

Following thematic tagging, group-level comparisons were conducted using frequency tables and Fisher's exact tests to assess whether theme prevalence differed significantly between NSI and VSI participants. This was supplemented by calculation-of-odds ratios and confidence intervals to support interpretability.

The second phase of analysis focused on responses to Q2 ("How would you describe the sound of an EV?"), which offered insight into participants' emotional and perceptual framing of EV acoustics. While Q1 explored broad conceptual differences between EVs and ICE vehicles, Q2 responses were more introspective and affective in nature. Each participant's Q2 response was analyzed to produce an average polarity score, ranging from –1 (strongly negative sentiment) to +1 (strongly positive sentiment). This continuous sentiment measure allowed for richer within-group variation to be captured, compared to simply binary-theme flags. Group-level summary statistics (i.e., mean, median, standard deviation, and proportions of negative, neutral, and positive responses) were calculated separately for NSI and VSI groups.

To assess whether the two groups differed in their sentiment distribution, a Wilcoxon rank-sum test was performed, which is a non-parametric alternative to the t-test more suitable for relatively small and potentially non-normally distributed samples such as in EVA. In addition, Cliff's delta was computed to quantify effect size and interpret practical significance. This combination of descriptive and inferential techniques allowed us to determine not just whether the groups differed, but also whether such differences had meaningful interpretive weight.

TABLE VI. OPEN-ENDED SURVEY QUESTIONS AND THEIR CORRESPONDING RESEARCH OBJECTIVES.

Code	Text	Research Purpose
Q1	How do you think EVs are different from other cars that use petrol or diesel?	To explore participants' baseline mental models of EV versus ICE vehicles. This provides insight into what characteristics (e.g., noise, environmental impact, performance) are most salient in shaping public attitudes.
Q2	How would you describe the sound of an EV?	To collect qualitative descriptors of EV acoustics, including specific features such as pitch, volume, and tonal quality. Responses provide insight into how these auditory features influence perceptions of comfort, safety, and detectability.
Q3	Have you ever had an experience where the sound of an electric car surprised you or caught you off guard?	To identify real-world examples of surprise or near-miss incidents, particularly those linked to detectability limitations. These responses offer a window into safety-related experiences and the emotional salience of such events.
Q4	What specific sounds or noises from electric cars do you find pleasant or unpleasant? Please describe them and explain why you feel that way.	To explore individual preferences and aversions related to EV sound characteristics. These insights help reveal what acoustic design elements may contribute to positive versus negative user experiences.

F. Analysis Methods for Open-Ended Questions 3 and 4

Questions 3 and 4 were designed to elicit more contextualized insights into participants' experiences and emotional reactions to EV sounds. To analyze these responses, we adopted a mixed approach combining response rate comparisons, nonparametric inferential tests, and polarity-based sentiment scoring.

For Question 3, which asked participants whether they had ever been surprised or caught off guard by an EV, analysis began with a response rate comparison between the NSI and VSI groups. A binary indicator was created for each participant reflecting whether they provided a valid response to Q3. We used a Fisher's exact test to assess whether response rates differed significantly between groups, reflecting potential differences in lived experience or salience of EV-related incidents.

Next, a sentiment analysis was conducted on the submitted free-text narratives, again using the *sentimentr* package in R. Only those participants who submitted non-blank responses were included in this phase. These per-person polarity scores enabled group-level comparisons of the affective tone of reported experiences. We employed a Wilcoxon rank-sum test to consider group differences (if any) in sentiment scores, with Cliff's delta reported as an effect size estimate. This combined approach once again allowed us to examine both how often participants described surprise-by-silence events and how emotionally valenced those reports were.

For Question 4, which asked participants to describe specific EV sounds they found pleasant or unpleasant, we again began by comparing response rates between groups using a Fisher's exact test. This assessed whether VSI participants were more likely than NSI to volunteer their views on EV sound characteristics. As with Q3, we created a binary flag for each participant indicating whether a non-blank response was recorded.

We then applied sentiment analysis to the valid Q4 responses using the same *sentimentr* pipeline. Each participant's description of pleasant or unpleasant EV sounds was scored for polarity, enabling quantitative comparison of sentiment between the two groups. Again, we used the Wilcoxon rank-sum test to assess statistical differences, with Cliff's delta providing a nonparametric measure of effect size. While thematic coding was not applied to Q4 responses in this phase, qualitative inspection of the comments supported interpretation of whether perceived pleasantness was driven by aesthetic preferences or by more functional concerns such as detectability.

Together, these analytical steps for Q3 and Q4 aimed to reveal whether group-level differences in experience, concern, or emotional response were evident; not just in how often participants commented, but in how they characterized and emotionally evaluated those experiences. This allowed for a deeper understanding of how EV sound design intersects with real-world safety, accessibility, and affective perception, particularly for vision-impaired pedestrians.

IV. RESULTS: DESCRIPTIVE ANALYSIS OF LIKERT ITEMS

This section presents the findings from qualitative analyses of the survey data. Results are organized thematically, beginning with the statistical comparison of Likert-scale items between sighted and vision-impaired participants, followed by multivariate analyses of response patterns. Section VII explores qualitative insights drawn from open-ended responses, providing a complementary perspective on participant experiences and perceptions of EV sounds.

A. Median Differences between NSI and VSI

Table VII presents the medians of both NSI and VSI groups across all nine Likert statements. From these results, vision-impaired participants have experienced EVs differently compared to sighted pedestrians.

TABLE VII. MEDIAN RATINGS FOR EACH LIKERT STATEMENT.

Statement (Condensed)	NSI	VSI
L1: I feel safe when I think there is an EV close by.	3	1
L2: It is easy to notice an EV approaching due to its sound.	2	1
L3: Sounds made by EVs help me understand what the vehicle is doing.	2	1
L4: I feel confident I understand an EV's next action based on its sound.	1	1
L5: I can react quickly to the sound of an EV when necessary.	3	1
L6: I find the sound of EVs pleasant.	4	2
L7: It takes little effort for me to listen to an EV's sound and understand what it is doing.	3	1
L8: I believe that the sound from all electric cars will be a positive thing for noise levels in busy cities and towns.	4	3
L9: Do you think it would be easy or hard to know when it is safe to cross the road in a busy EV environment?	3	2

The VSI group consistently reported lower median ratings across all nine Likert-scale items, suggesting a pronounced difference in how pedestrians with vision-impairments perceive and interpret the sounds of EVs compared to sighted participants. In terms of perceived safety (L1), VSI participants expressed markedly less confidence, with a median rating of 1, compared to a neutral 3 in the NSI group. Detectability of EVs by sound (L2, L3) also emerged as a key area of divergence, with VSI participants again rating 1, indicating significant challenges in perceiving or interpreting EV acoustics. This does not suggest that the VSI Group has any auditory deficit compared to the NSI Group, rather, the difference is more likely explained by the sensory context – as in, NSI participants can confirm the presence or absence of an EV through sight, whereas VSI participants must rely solely on auditory cues.

Both groups expressed low confidence in predicting vehicle intent (L4), with medians of 1, pointing potentially to a fundamental design flaw in AVAS signals, which may not communicate directional or behavioral intent effectively. Reaction time (L5) revealed another significant contrast, with NSI respondents rating 3 (neutral) and VSI rating 1, suggesting that participants with vision-impairments feel substantially less prepared to respond quickly to approaching EVs based on sound alone.

NSI participants found EV sounds more pleasant overall (L6) and were more optimistic about their impact on urban noise environments (L8). Conversely, VSI participants appeared more critical or uncertain in both respects, suggesting different acoustic and safety needs. Finally, in L9, which addressed the complexity of real-world street navigation amid multiple EVs, VSI respondents gave a median of 2 (more difficult), while NSI participants remained neutral (3). This further reinforces that current AVAS implementations may not be meeting the navigational needs of vision-impaired pedestrians in dynamic environments.

B. *Interquartile Ranges for NSI and VSI Groups*

The IQR values reveal important differences in how consistently each group responded to the survey items. The VSI group showed lower variability across most items, with five of the nine Likert items showing an IQR of 0, indicating that at least half of the vision-impaired respondents gave the same or very similar responses (see Table VIII). This suggests high agreement or shared perception within this group, particularly in items L2, L3, L4 and L7; all of which deal with detectability and interpretability of EV sounds.

In contrast, the NSI group displayed greater dispersion in their responses, with IQRs ranging from 1 to 3. Notably, items such as L2 and L7 had IQRs of 3, reflecting substantial variability in how sighted participants perceive the usefulness and clarity of EV sounds. This may reflect the fact that sighted individuals rely less consistently on auditory cues, leading to broader variation in their experiences and interpretations. Interestingly, both groups shared a moderate degree of variability in response to the sound pleasantness item (L6) and urban noise attitudes (L8), suggesting more subjective or context-dependent opinions across cohorts on these dimensions. These IQRs highlight that whilst VSI participants tended to rate the AVAS signals more negatively, they also did so with greater internal consistency, indicating a shared perceptual experience.

TABLE VIII. INTERQUARTILE RANGE (IQR) FOR EACH LIKERT ITEM BY PARTICIPANT GROUP.

Code	NSI	VSI
L1	2.25	1
L2	3	0
L3	2	0
L4	2	0
L5	2	1
L6	2	2
L7	3	0
L8	2	3
L9	1	1

C. *Overview of Group Response Density*

To complement the median and IQR data, violin plots incorporating a kernel density estimate were generated to visualize the underlying shape and distribution of responses across both groups (see Figure 2). These plots highlight not only central tendencies but also the frequency and symmetry of responses, revealing several notable patterns across Likert statements for each group. Across most items (L1 to L8), the NSI group demonstrated a broader spread of responses and more positively skewed distributions than the VSI group. For instance, statements L1, L2, and L5 show a visibly higher concentration of NSI responses in the upper end of the scale, with little to no overlap with VSI distributions, suggesting strong perceptual divergence. In contrast, items like L6 and L8 show slightly more overlap, although a positive bias remains evident in the NSI group.

Particularly notable are the relatively narrow and consistent response patterns within the VSI group across statements L2 to L4 and L7, where responses cluster tightly at the lowest scale values. This reinforces earlier descriptive findings (medians and IQRs) and suggests a high degree of consensus in negative perceptions within the VSI cohort for these statements.

D. *Estimating Magnitude of Difference*

While distribution plots and descriptive measures such as medians and IQRs offer insight into general response patterns, they do not quantify per se the magnitude of difference between groups. To address this, effect size statistics (δ) were calculated for each Likert item to estimate the strength of association between group membership (NSI vs VSI) and response level.

Importantly, effect size also serves a complementary role to significance testing. While p-values test whether an observed difference is likely due to chance, they are sensitive to sample size and do not convey the magnitude of that difference. However, a large effect size accompanied by a statistically significant p-value provides strong evidence that the group difference is both reliable and meaningful in practical terms.

The effect size results (see Table IX) reveal substantial group-level divergence between NSI and VSI participants in their responses to almost all survey Likert items. Notably, seven out of nine statements yielded large effect sizes ($|\delta| > 0.474$), indicating that participants in the VSI group were consistently more likely to rate items negatively and the NSI group more positively; a pattern that points to systematic differences in experience with EVs.

L1 and L2, which focus on the general detectability and clarity of EV detection, show the largest effect sizes ($|\delta| = 0.619$ and 0.617, respectively). This suggests that NSI participants may be more confident or comfortable with EVs in daily contexts, potentially because NSI pedestrians rely less on auditory cues for safety and spatial awareness.

L5 to L8 also register large effect sizes, reflecting perceptual dimensions such as ease of localization, emotional comfort, and confidence when crossing roads near the presence of EVs. This suggests that the auditory features of EVs may be less salient than intended, insufficient on their own to enable consistent detection, and ultimately require visual confirmation to ensure safe navigation. This ambiguity may play a role in VSI pedestrians' degree of trust (or lack thereof) in the acoustic cues embodied in AVAS.

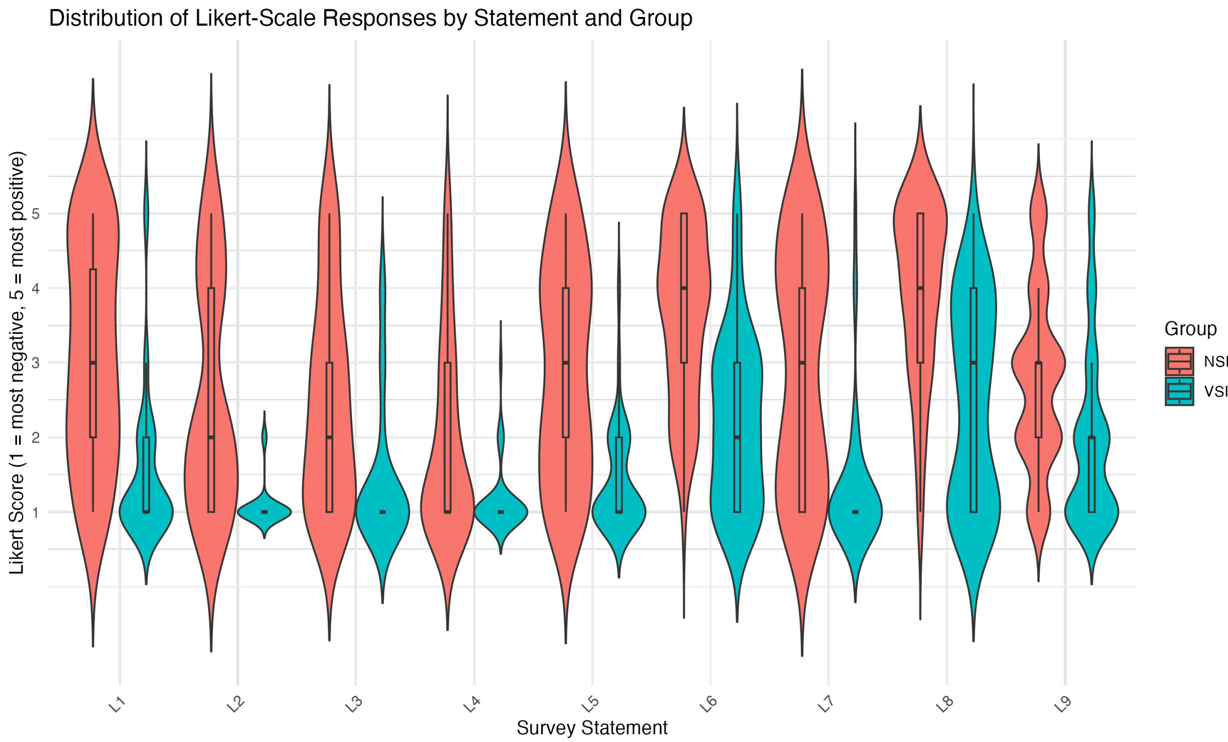

Figure 2. Violin plots of Likert-scale responses across nine statements, grouped by NSI (red) and VSI (blue). Each plot shows response density, with embedded boxplots indicating median and interquartile range (IQR). Higher Y-axis values mean more positive responses, broader X-axis ranges mean more within-group consensus.

TABLE IX. CLIFF'S DELTA (δ) EFFECT SIZE ESTIMATES FOR EACH LIKERT-SCALE ITEM COMPARING NSI AND VSI GROUPS.

Code	Delta	Magnitude
L1	0.619	Large
L2	0.617	Large
L3	0.46	Medium
L4	0.36	Medium
L5	0.589	Large
L6	0.602	Large
L7	0.589	Large
L8	0.584	Large
L9	0.404	Medium

The medium effect sizes observed in L3, L4, and L9 suggest more modest but still meaningful divergences. For example, L3 and L4, which capture aspects of perceived risk or uncertainty, may reflect more subjective or situational variability, even among VSI participants. The medium effect in L9, which concerns overall perceived safety in near-future environments occupied by many EVs, points to a nuanced area where both groups share some common ground.

Taken together, these effect-size magnitudes suggest that the acoustic characteristics and perceived behavioral implications of EV sounds are experienced quite differently depending on if a pedestrian has a vision sensory impairment or not. The consistently high δ values in favor of the NSI group suggest that VSI participants are not receiving clear or unambiguous auditory cues from the AVAS systems used in EVs, highlighting a potential gap in the effectiveness of these signals for vision-impaired pedestrians.

V. Results: Correlation Analysis on likert items

Spearman's rank correlation was employed to quantify the strength and direction of monotonic associations between pairs of Likert-scale statements. This non-parametric method allowed us to explore whether higher agreement with one statement corresponded with higher (or lower) agreement with another. Separate correlation matrices were generated for the NSI and VSI groups to examine how patterns of internal consistency, alignment, or divergence differed within each participant cohort. Table X presents Spearman's ρ (rho) correlation coefficients for the NSI group, while Table XI displays the corresponding values for the VSI group.

Figure 3 visualizes the Spearman's rank correlation coefficients between all pairs of Likert items for the NSI group using a heatmap. The darker blue shades represent stronger positive correlations, while lighter shades indicate weaker associations. This matrix reveals several strong and consistent correlations among items related to auditory awareness, detectability, and clarity (e.g., L1–L5, L1–L6, L2–L3), suggesting that participants without sensory impairments formed coherent impressions of the AVAS stimuli. The strong correlations indicate that when one aspect of AVAS performance (e.g., sound detection) was rated highly, other related aspects (e.g., sound clarity or usefulness) were also likely to receive high ratings. Conversely, minimal correlations involving L7–L9 suggest more varied responses to statements relating to real-world decision-making or ambiguity, indicating greater diversity of interpretation in those areas.

Figure 4 presents the Spearman's rank correlation matrix for the VSI group. Compared to the NSI group, stronger correlations are concentrated around statements L3 to L5, indicating that participants with vision-impairments tended to show higher internal consistency when rating items related to clarity, usefulness, and environmental context. The very strong relationship between L3 and L4 suggests a shared perceptual framing of these statements; likely rooted in lived-experience that calls for navigating environments without visual information. Correlations across other item pairs are generally lower or more varied, reflecting greater diversity in how individual VSI participants interpreted or experienced AVAS cues across dimensions such as detection (L1) and confidence (L9). This pattern may point to variability in how effective current AVAS implementations are at delivering consistent or meaningful messaging for vision-impaired users.

TABLE X. Spearman's ρ correlation matrix for the NSI group, showing strength and direction of relationships between responses to different Likert-scale statements

	L1	L2	L3	L4	L5	L6	L7	L8	L9
L1	1.00	0.53	0.47	0.53	0.58	0.62	0.07	0.29	0.51
L2	0.53	1.00	0.55	0.43	0.52	0.27	0.11	0.14	0.36
L3	0.47	0.55	1.00	0.54	0.53	0.23	0.17	0.38	0.37
L4	0.53	0.43	0.54	1.00	0.53	0.30	0.20	0.24	0.27
L5	0.58	0.52	0.53	0.53	1.00	0.31	0.22	0.18	0.49
L6	0.62	0.27	0.23	0.30	0.31	1.00	0.17	0.43	0.20
L7	0.07	0.11	0.17	0.20	0.22	0.17	1.00	0.03	- 0.01
L8	0.29	0.14	0.38	0.24	0.18	0.43	0.03	1.00	0.22
L9	0.51	0.36	0.37	0.27	0.49	0.20	- 0.01	0.22	1.00

TABLE XI. Spearman's ρ correlation matrix for the VSI group, indicating the strength and direction of relationships between responses to each Likert-scale statement.

	L1	L2	L3	L4	L5	L6	L7	L8	L9
L1	1.00	0.16	0.21	0.31	0.35	0.17	0.27	0.13	-0.03
L2	0.16	1.00	0.32	0.45	0.37	0.16	0.10	0.26	-0.01
L3	0.21	0.32	1.00	0.74	0.43	0.49	0.34	0.31	0.40
L4	0.31	0.45	0.74	1.00	0.55	0.32	0.38	0.34	0.30
L5	0.35	0.37	0.43	0.55	1.00	0.24	0.31	0.41	0.33
L6	0.17	0.16	0.49	0.32	0.24	1.00	0.19	0.46	0.28
L7	0.27	0.10	0.34	0.38	0.31	0.19	1.00	0.39	0.32
L8	0.13	0.26	0.31	0.34	0.41	0.46	0.39	1.00	0.22
L9	-0.03	-0.01	0.40	0.30	0.33	0.28	0.32	0.22	1.00

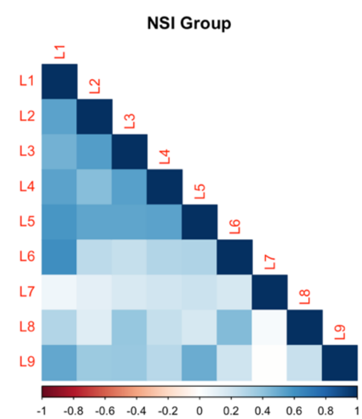

Figure 3. Spearman's rank correlation heatmap for the NSI group. Each cell shows the strength of monotonic association between pairs of Likert-scale statements, with color intensity representing correlation strength (ρ). Stronger positive correlations are indicated by darker blue. The pattern suggests coherent perceptions of AVAS features among NSI participants, particularly across L1–L6.

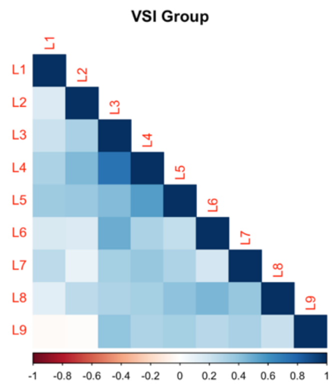

Figure 4. Spearman's rank correlation heatmap for the VSI group. Strong positive correlations are most prominent between Likert items L3–L5, suggesting coherent responses around statements relating to AVAS clarity and usefulness. Broader variation in correlation strength indicates more heterogeneous interpretation across other perceptual dimensions compared to the NSI group.

VI. RESULTS: SIGNIFICANCE TESTING ON LIKERT ITEMS

To assess whether the NSI and VSI groups differed significantly in their responses to each AVAS-related Likert statement, a series of Mann-Whitney U tests were conducted. For each item, the U statistic and p-value were computed to determine the statistical significance of group differences. In addition, the Rank-Biserial Correlation (r) was calculated to provide a measure of effect size, offering insight into the practical magnitude of any observed differences. The results are presented in Table XII, highlighting both statistical and practical significance across the full set of statements.

This was followed by a multivariate analysis (PERMANOVA) to determine whether the overall response patterns across all Likert statements differed significantly between NSI and VSI groups. A Gower distance matrix was employed with 1 degree of freedom (df) as group difference was the single dimension under consideration. Spearman's rank correlation was employed to quantify the strength and direction of monotonic associations between pairs of Likert-scale statements.

A. Group Differences in Likert Responses

Table XII presents the results of Mann-Whitney U tests (U), p-values, Z scores, and rank-biserial correlation coefficients (r) as measures of effect size. All nine Likert items yielded statistically significant differences ($p < 0.001$) between the NSI and VSI groups. Of these, six items L1 (detectability of EVs by sound); L2 (clarity of EV approach); L5 (sound helped judge vehicle location); L6 (sound helped judge vehicle movement); L7 (sound gave enough warning); and L8 (sound helped feel safe) showed large effect sizes ($r \geq 0.5$), indicating substantial differences in perceptual experience between the groups. Items L3 (awareness of vehicle presence), L4 (ability to detect direction), and L9 (need for visual confirmation) demonstrated medium effect sizes, suggesting meaningful but more moderate group differences.

The strongest effects were observed for L2 ($r = 0.599$) and L1 ($r = 0.552$), highlighting that VSI participants were significantly less likely than their NSI counterparts to find the sound of EVs sufficiently clear or noticeable. This most likely reflects not a sensory deficit on the part of VSI respondents, but rather that NSI participants can supplement or bypass auditory ambiguity through visual verification, whereas those with vision-impairments are more vulnerable to the limitations in AVAS design.

Similarly, consistent large effects across L5 to L8 suggest that NSI pedestrians may not depend on AVAS to convey spatial or safety-related cues, while VSI participants (who must rely on these sounds) perceive them as falling short of their intended function. In this sense, the findings do not imply that VSI respondents struggle more with sound itself, but rather that the current auditory cues are not fit for purpose when visual compensation is not possible. These findings point to a systematic divergence in experience between groups, which is while AVAS may be minimally functional for those with full vision, it does not adequately support the needs of vision-impaired pedestrians. The consistently large and medium effect sizes, coupled with highly significant p-values, confirm that these differences are both statistically robust and practically meaningful.

TABLE XII. RESULTS OF MANN–WHITNEY U TESTS COMPARING RESPONSES TO EACH LIKERT STATEMENT BETWEEN NSI AND VSI GROUPS.

Statement	U	p-value	Z	r	Magnitude
L1	1516	<0.001	5.15	0.552	Large
L2	1514	<0.001	5.59	0.599	Large
L3	1366	<0.001	4.02	0.431	Medium
L4	1272	<0.001	3.51	0.376	Medium
L5	1487	<0.001	4.95	0.530	Large
L6	1500	<0.001	4.91	0.526	Large
L7	1487	<0.001	5.09	0.546	Large
L8	1482	<0.001	4.78	0.513	Large
L9	1314	<0.001	3.33	0.357	Medium

B. Group Difference in Overall Response Pattern

To assess whether the overall patterns of Likert-scale responses differed significantly between groups, a Permutational Multivariate Analysis of Variance (PERMANOVA) was conducted using a Gower distance matrix and 999 permutations. This non-parametric analysis tests whether overall response patterns across the nine statements differ by group (NSI vs. VSI). As shown in Table XIII, the PERMANOVA revealed a statistically significant group difference ($p = 0.001$), with the grouping factor (NSI vs. VSI) explaining approximately 30.6% of the variance in response patterns ($R^2 = 0.30583$). This indicates a meaningful divergence in how the two groups rated the statements when considered collectively, not just item by item.

A Principal Coordinates Analysis (PCoA) visualization (see Figure 5) based on the Gower distance matrix further illustrates this divergence, with a clear separation between NSI and VSI participants in the multivariate response space. The elliptical contours around each group represent 95% confidence intervals, reinforcing the systematic nature of this separation.

These findings therefore reinforce the earlier Mann–Whitney U test results by showing that the divergence between NSI and VSI groups is not only significant at the individual Likert-statement level, but also consistent and substantial when evaluating the overall configuration of their responses.

A remaining 69.4% of variance remains unexplained and this residual variation may reflect individual differences unrelated to sensory status - such as prior exposure to EVs, age, mobility habits, mobility routes, cognitive strategies, or situational context during sound encounters (e.g., background noise, urban design, time of day). These factors, while outside the scope of the present analysis, merit future exploration to further characterize how pedestrians interpret AVAS cues.

TABLE XIII. PERMANOVA RESULTS COMPARING OVERALL RESPONSE PATTERNS BETWEEN NSI AND VSI GROUPS. PERMANOVA BASED ON GOWER DISTANCE WITH 999 PERMUTATIONS.

Source	df	Sum of Squares	R^2	F	p-value
Group	1	2.0421	0.30583	37.448	<0.001
Residual	85	4.6353	0.69417		
Total	86	6.6774	1.00000		

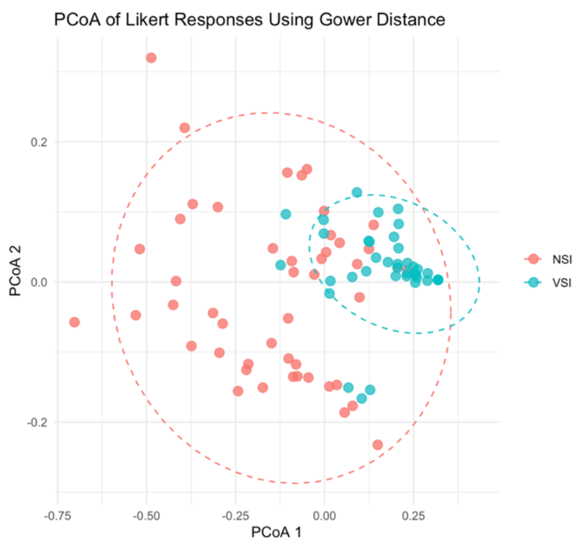

Figure 5. PCoA plot based on Gower distances calculated from Likert-scale responses (L1–L9). Each point represents an individual respondent, coloured by group. Ellipses indicate 95% confidence intervals for each group. The clear separation of the VSI cluster supports the significant group difference detected in the PERMANOVA ($R^2 = 0.306$, $p < .001$).

VII. RESULTS: ANALYSIS OF OPEN-ENDED RESPONSES

A. Analysis of Question 1

Thematic coding of Question 1 responses revealed that *Audibility* was the only theme to exhibit a statistically significant difference in prevalence between NSI and VSI participants. The odds of *Audibility*-related references were over ten times higher among one group compared to the other (OR = 10.21, 95% CI [1.25, 475.92], $p = 0.013$). This marked effect contrasts with the remaining five themes, for which no statistically reliable group differences were observed. Although *Comfort* and *Environment* featured moderately in both groups, the associated odds ratios hovered close to unity and confidence intervals were wide, indicating negligible and uncertain group-level variation. *Cost* and *Performance* were comparatively infrequent, and *Safety* was rarely mentioned at all, producing unstable estimates with correspondingly broad confidence bounds.

When considered together, these results suggest that perceptions linked to *Audibility* form the primary thematic distinction between NSI and VSI respondents when articulating how EVs differ from ICE vehicles (Q1), whereas other evaluative dimensions appear more uniformly distributed across groups (see Figure 6).

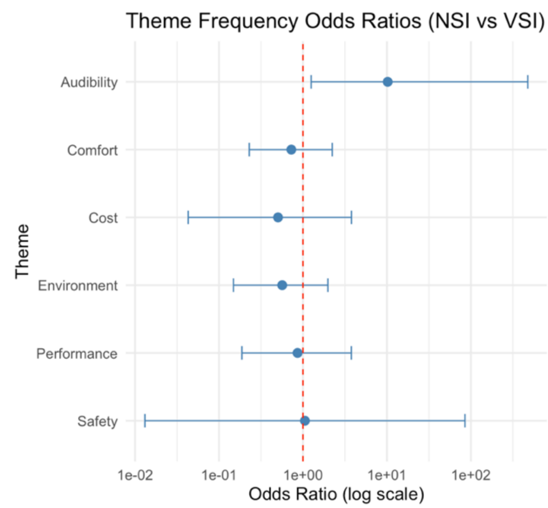

Figure 6. Theme frequency odds ratios (NSI vs VSI) for Q1. The plot shows odds ratios with 95% confidence intervals on a log scale.

B. Analysis of Question 2

A polarity-based sentiment analysis was conducted on participants' open-text responses to "How would you describe the sound of an EV?" This approach assigns each response a sentiment score ranging from –1 (entirely negative tone) to +1 (entirely positive tone), with scores near zero indicating a neutral tone. The analysis found that mean sentiment was positive for both groups, with NSI respondents showing a slightly higher average sentiment (M = 0.24, SD = 0.68) compared to VSI respondents (M = 0.16, SD = 0.64). These values suggest that, on average, both groups tended to use language with a mildly positive valence when describing EV sounds themselves, although variability within each group was also high.

C. Analysis of Question 3

Sentiment analysis revealed notable differences in the emotional tone of surprise experiences between groups (see Table XIV).

VSI participants demonstrated greater emotional variability in their surprise experiences (SD = 0.216 vs 0.121) and were three times more likely to report negative sentiment (31.6% vs 10.3% negative responses). While both groups clustered around neutral sentiment overall, the broader range and higher frequency of negative responses among VSI participants suggests these surprise-encounters represent genuine safety concerns rather than neutral observations. The polarized nature of VSI responses (ranging from highly negative (-0.288) to highly positive (0.8)), indicates more emotionally charged experiences, consistent with the safety implications of reduced EV detectability for vision-impaired pedestrians.

However, while the descriptive statistics above suggest VSI participants tend to experience more negative sentiment, the formal statistical test fails to confirm this pattern (see Figure 7 where $p = 0.104$). In addition, the small effect size ($\delta = 0.215$) should also be interpreted cautiously, as it may reflect sampling variability rather than a true group difference. With 77 total responses, the study had a reasonable sample size relative to how many people tend to respond to such niche surveys, so there was a relatively good chance that meaningful differences would be detected in this question if they existed. The non-significant result therefore suggests that both groups experience similar emotional responses to EV surprise encounters, regardless of sensory status; and indicates that while VSI participants may face greater practical safety challenges with EV detection, their emotional reactions to surprise incidents are not systematically different from those of NSI participants. Note that this similarity in emotional impact likely does not validate current AVAS design but rather indicates that inadequate acoustic warning systems may be equally problematic for all pedestrians, eliciting comparable levels of distress across sensory abilities. Further research may be warranted, therefore, to develop acoustic cues that provide an advance "heads-up" alert on approach, helping to reduce the anxiety or discomfort associated with such encounters.

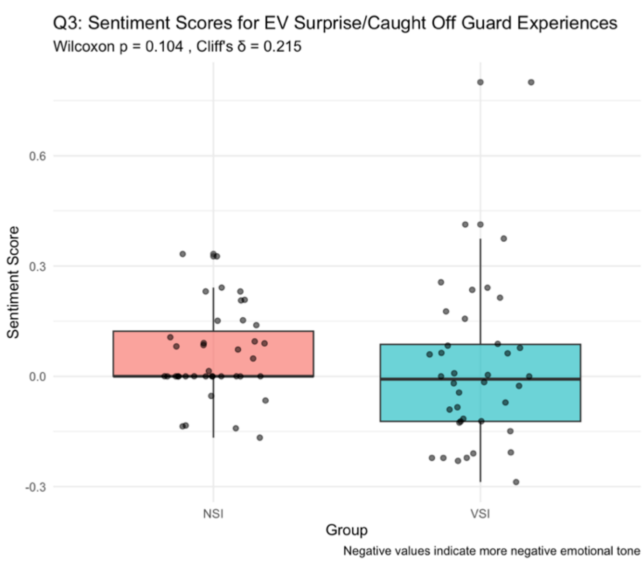

Figure 7. Sentiment scores for Q3 responses describing experiences of being surprised or caught off guard by electric vehicles, by participant group. Box plots show median, quartiles, and range; individual data points are jittered to show distribution.

D. Analysis of Question 4

An examination of Q4 revealed distinct patterns in the underlying motivations driving participants' evaluations of EV sounds (see Table XV), demonstrating a clear reversal in priorities between groups. NSI participants predominantly framed their responses around aesthetic preferences (50.0%), focusing on whether EV sounds were pleasant, soothing, or enjoyable. In contrast, VSI participants primarily emphasized functional concerns (51.5%), concentrating on detectability, warning capability, and safety implications. This pattern suggests that when evaluating EV acoustic characteristics, NSI participants approach the question from a comfort and pleasantness perspective, while VSI participants evaluate the same sounds through a safety and utility lens. Also, the higher proportion of mixed/unclear responses among VSI participants (33.3% vs 23.3% for NSI) may reflect the complex interplay between aesthetic judgment and practical necessity when auditory information is critical for navigation.

Indeed, follow-on sentiment analysis of Q4 responses revealed a statistically significant difference between groups when participants evaluated pleasant and unpleasant characteristics of EV sounds (Figure 8). NSI participants demonstrated significantly more positive sentiment (Wilcoxon $p = 0.024$, Cliff's $\delta = 0.331$, medium effect size) when describing EV acoustic features compared to VSI participants. This finding contrasts with the Q3 results, where no group differences emerged in emotional responses to surprise encounters. The significant difference in Q4 suggests that when explicitly evaluating EV sound characteristics, participants' sensory status may fundamentally influence their assessment. NSI participants, who predominantly focus on aesthetic qualities, tend to frame EV sounds more positively, likely emphasizing features such as quietness as pleasant or soothing. Conversely, VSI participants, whose evaluations center on functional concerns, express more negative sentiment, reflecting frustration with inadequate detectability.

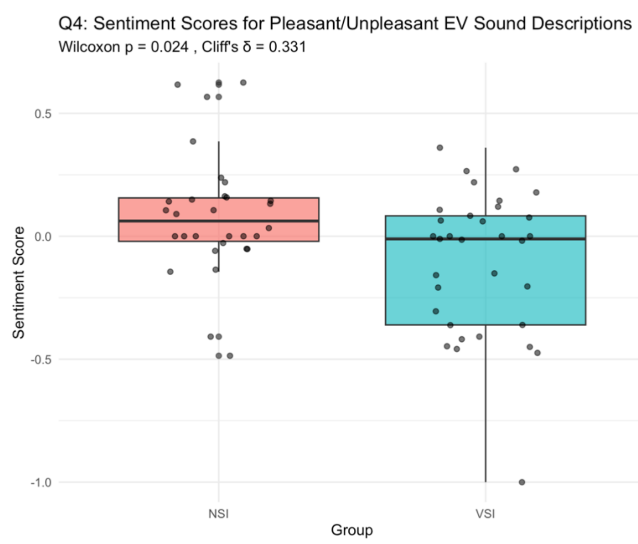

Figure 8. Sentiment scores for Q4 responses describing pleasant and unpleasant characteristics of EV sounds, by participant group.

TABLE XIV. PERMANOVA RESULTS COMPARING OVERALL RESPONSE PATTERNS BETWEEN NSI AND VSI GROUPS. PERMANOVA BASED ON GOWER DISTANCE WITH 999 PERMUTATIONS.

n	mean_sentiment	median_sentiment	sd_sentiment	min_sentiment	max_sentiment	negative_responses	neutral_responses
NSI (39)	0.05657037	0	0.12122559	-0.1668115	0.3328201	4	24
VSI (38)	0.01913296	-0.0078763	0.21590186	-0.2877128	0.8	12	17
				positive_responses	**%_negative**	**%_neutral**	**%_positive**
			NSI (39)	11	10.3	61.5	28.2
			VSI (38)	9	31.6	44.7	23.7

TABLE XV. PERMANOVA RESULTS COMPARING OVERALL RESPONSE PATTERNS BETWEEN NSI AND VSI GROUPS. PERMANOVA BASED ON GOWER DISTANCE WITH 999 PERMUTATIONS.

Group	Primary Concern	n	%
NSI	Aesthetic	15	50
NSI	Functional	8	26.7
NSI	Mixed/Unclear	7	23.3
VSI	Aesthetic	5	15.2
VSI	Functional	17	51.5
VSI	Mixed/Unclear	11	33.3

E. Summary of Open-Ended Results

The analysis of open-ended responses reveals an interesting progression of how sensory status may shape EV-related perceptions and experiences. VSI participants demonstrated fundamentally different baseline mental models of EVs, being ten times more likely to conceptualize them through an audibility framework (Q1), establishing distinct cognitive foundations between groups starting off. While both groups initially described EV sounds with similar mildly positive sentiment (Q2), their evaluative frameworks diverged sharply when explicitly assessing pleasant/unpleasant sound characteristics (Q4); with NSI participants focusing on aesthetic qualities and expressing more positive sentiment versus VSI participants emphasizing functional concerns and expressing significantly more negative evaluations.

Interestingly, both groups demonstrated comparable emotional distress when describing surprise encounters with EVs (Q3), suggesting that there may be AVAS inadequacies that create universal pedestrian distress in surprise scenarios rather than this being an VSI-specific issue. This pattern indicates that EVs fail both groups but for different reasons: VSI participants experience inadequate functional warning (possibly being distance independent), while NSI participants still encounter distressing surprise incidents that suggest insufficient acoustic range or volume to provide adequate advance warning at appropriate distances.

Collectively, these findings suggest that acoustic stimuli may be processed through fundamentally different perceptual and evaluative frameworks depending on sensory status, yet result in similar emotional distress across groups. This indicates that there may still be a need to further refine AVAS design for addressing both functional detectability requirements and adequate detection distances for universal pedestrian comfort. This challenges the assumption that current acoustic warning systems adequately serve any pedestrian population through either sound quality or propagation characteristics.

VIII. DISCUSSION

A. Notification Timing, Distance, and AVAS Design

The findings of the EVA study highlight a systematic divergence in how sighted and vision-impaired pedestrians perceive and respond to EV sounds and their associated AVAS. Across all nine Likert-scale statements, participants with vision-impairments consistently reported more negative perceptions of safety, detectability, and interpretability compared to sighted participants. This difference was not only statistically significant but also accompanied by large effect sizes in many items, indicating practical, real-world implications for pedestrian safety.

A key insight emerging from the study is the role of distance and notification timing in shaping pedestrian experience. Participants, particularly those with vision-impairments, reported difficulties in detecting EVs early enough to allow safe and confident reactions. Surprise encounters were common across both groups, suggesting that the current design of AVAS may not provide adequate advance warning on a universal level. While regulations specify minimum sound levels and frequency characteristics, they do not explicitly address how far in advance an EV should be detectable or whether pedestrians should be given a clear “heads-up” alert to reduce surprise. In essence, distance is operationalized only in terms of measurement geometry and not in terms of real-world perceptual outcomes. This regulatory vagueness leaves a critical gap: where compliance testing ensures only that vehicles emit sound at measurable levels but has no guidance on how that sound should function with distance in real urban environments to ensure meaningful detectability for all users.

The EVA findings on late or surprise detection of EVs resonate strongly with concerns already documented during UNECE's early QRTV sessions. For example, German BASt trials in 2011 [19] demonstrated that some ICE comparators - even those used in other submitted trials as benchmarks - were effectively as quiet as EVs at low speeds, raising questions about whether "quietness" should be treated as an EV-only issue. EVA participant-reports of delayed detection and reliance on visual cues suggest that this ambiguity persists in real-world pedestrian experience, and that the current minimum thresholds specified in UNECE R138 do not sufficiently account for this overlap. Similarly, EVA survey responses describing unexpected encounters with smaller micromobility devices echo an omission in the regulations: despite early debate during 2010-2011 about whether e-bikes or powered two-wheelers should be included, the scope was ultimately restricted to M and N categories (passenger cars and light goods vehicles). Considered together, the EVA results confirm that unresolved issues identified in UNECE's formative years remain active sources of risk and uncertainty in today's pedestrian experience more than a decade later.

Indeed, sounds may meet regulatory thresholds at close range on controlled test-tracks, but may fail to propagate effectively in dynamic, noisy environments where early detection is essential. Recent experimental work supports this distinction between detectability and functional perceptual efficiency. For example, Müller, Forssén, and Kropp [65] found that although AVAS signals can be detectable in controlled conditions, localization accuracy deteriorates significantly, particularly for tonal signals and worsens further when multiple vehicles are present. This highlights a limitation in standards whereby they equate threshold detectability with effective pedestrian awareness.

Moreover, while the updated ISO 16254 standard introduces more sophisticated measurement techniques with expanded microphone arrays and more refined psychoacoustic metrics, it still frames AVAS evaluation in terms of signal production and measurement, rather than alert effectiveness as a function of distance, pedestrian experience, or real-world noisy environments. Many academic studies have largely mirrored this focus, emphasizing laboratory-based detection thresholds and signal characteristics, but relatively little work has been done on how AVAS design could be optimized to improve anticipatory awareness and reduce surprise under real-world conditions. The forthcoming amendments in R138.02, which explicitly integrate these ISO updates into the regulatory framework, reflect this same trajectory: the regulation evolves by refining *how* signals are measured (e.g., multi-microphone arrays, clearer specifications for reverse sound) without substantively expanding *what* is measured. This iterative loop between UNECE regulation and ISO procedure ensures technical consistency but risks reinforcing a narrow emphasis on compliance metrics, rather than addressing the broader perceptual challenges consistently highlighted in end-user studies such as EVA.

B. Functional Meaning versus Branded AVAS

Another critical insight from the EVA study concerns the potential lack of functional meaning in current AVAS implementations. While EVs are required to emit sound that varies with speed, the standards require relatively coarse tonal features, leaving substantial latitude for car manufacturers to design uniquely branded sound-signatures. This flexibility has led to the emergence of distinctive acoustic identities for EVs, often emphasizing marketing differentiation over communicative function. Although such sounds can pass regulatory compliance tests, they may be failing to provide pedestrians with actionable cues about vehicle state or even driver intent as ICE vehicles often inherently do.

The survey results suggest that vision-impaired participants in particular seek sounds that are informative rather than aesthetic. Their framing of EV sounds as tools for detection and orientation contrasts with the sighted participants' emphasis on sound quality or pleasantness. This divergence highlights a tension in AVAS design: while branding may enhance product identity and user-experience for drivers, it risks undermining the functional purpose of AVAS for pedestrians who depend on it.

In terms of functional cues, ICE vehicles provide a richer palette of acoustic information than simply indicating speed, reverse, or stationary statuses. While these sounds may not always be aesthetically pleasing, they are closely tied to the mechanical operations of the engine. Subtle variations in engine tone can signal acceleration or deceleration, turning intent, driver impatience, even driver experience-level, as well as conditions such as engine misalignment on loose terrain or maneuvering in constrained spaces. Such cues give pedestrians valuable anticipatory information, helping them navigate traffic environments more safely and confidently. By contrast, current AVAS implementations often reduce this communicative richness to uniform, synthetic tones that satisfy baseline detectability requirements but lack the dynamic subtleties needed to support true situational awareness.

Future research should therefore extend beyond compliance-based assessments to examine how ICE soundscapes can inform the development of AVAS with functional meaning. Such studies could identify what acoustic dimensions (such as timbral shifts, rhythmic variations, or dynamic modulation) are most effective for conveying intent. By embedding these subtle yet informative cues into AVAS, systems could evolve from merely meeting regulatory requirements into genuinely communicative tools, enhancing not only detection but also pedestrians' ability to interpret and anticipate vehicle-driver behavior.

C. Future Urban Soundscapes

A further consideration emerging from the EVA study relates to the future soundscape of urban environments as EVs proliferate. If all slow-moving traffic in a busy street comprises EVs with each emitting similar synthetic AVAS tones, pedestrians may face new challenges in interpreting and navigating this new acoustic environment. While individual AVAS designs can be assessed in isolation for detectability and compliance, the collective effect of multiple overlapping signals has not yet been meaningfully examined. AVAS tones that are designed to be individually detectable may, when layered together in dense traffic, create a homogenized and fatiguing auditory environment - ironically echoing the very concerns about noise pollution that traffic reduction campaigns have sought to address for decades in relation to ICE vehicles. Rather than enhancing safety, such soundscapes could overwhelm pedestrians with undifferentiated stimuli, reducing their ability to localize specific vehicles or interpret their behavior. The relative uniformity of AVAS (typically composed of tonal structures tied to speed) raises questions of differentiation. Will pedestrians be able to distinguish between vehicles approaching from different directions, or will multiple AVAS signals blur into a single background hum?

From the perspective of urban residents, these issues extend beyond safety into quality of life. Current regulatory frameworks are silent on the broader implications of large-scale AVAS deployment, focusing on minimum requirements for audibility without addressing the cumulative impact on the urban soundscape. If synthetic tones dominate city centers, there is a risk of creating a new form of noise pollution, one that is both uniform and omnipresent.

The challenge, therefore, is not only to ensure that AVAS provide sufficient detectability in isolation but also to consider their collective ecological validity in real-world environments. Solutions may involve introducing richer acoustic palettes that allow vehicles to remain distinct from one another or exploring adaptive sound designs that respond to environmental context (e.g., adjusting timbre or modulation in crowded versus quiet streets). Again, Müller, Forssén, and Kropp [65] provide empirical evidence of this problem occurring in current AVAS when two or three vehicles employed the same or similar tonal signal. Their research shows that participants frequently failed to localize individual AVAS sources, with tonal designs performing worst. Their findings reinforce concerns that uniform AVAS soundscapes in busy urban contexts may not only blur together but actively impair pedestrians' ability to navigate safely. Interestingly, their research demonstrates that ICE vehicles are more robust in these scenarios.

D. Bridging the Research–Standards Gap in AVAS

Despite the immediate safety implications of AVAS, there remains a striking gap between academic research and the development of standards, policy, and regulation. To date, AVAS standards have been primarily industry-driven, with efforts focused on measurable compliance parameters such as minimum sound levels and frequency components. While this has ensured regulatory clarity for manufacturers, it has left broader questions around perceptual effectiveness, functional meaning, and real-world pedestrian experience largely unaddressed. It is important to acknowledge that UNECE's QRTV process did invite and receive extensive submissions, which generated a substantial body of research data. However, much of this evidence came from third-party contributors, including a strong presence from industry and affiliated research groups. By contrast, the U.S. NHTSA directly commissioned and conducted structured human trials between 2011 and 2013, providing perhaps a more authoritative baseline for how pedestrians detect and interpret vehicle sounds.

UNECE's reliance on submitted data meant that the scope, comparators, and methodologies varied widely across trials, and it must be said that in some cases lacked transparency about critical parameters such as what ICE vehicles were used as benchmarks. Both processes ultimately converged on similar regulatory compromises, prioritizing minimum SPL and frequency thresholds over perceptual outcomes such as localization accuracy, notification distance, or anticipatory awareness. This institutional pathway is significant: it seems that authority-led trials yielded stronger and more consistent evidence, while UNECE's submissions-based model introduced variability that often diluted end-user perspectives. In this context, the likes of the EVA study may help to fill a gap in Europe by providing independent, systematically gathered end-user evidence that speaks directly to the perceptual challenges that AVAS regulations have struggled to resolve. Moreover, because ISO standards are a mechanism through which UNECE requirements are engaged and operationalized, the rigorous ISO and CEN standards processes may provide the most effective channel for incorporating academic research into the regulatory framework in future revisions. Yet, the opportunity to embed such evidence more systematically into the standards process remains limited, reflecting a broader structural disconnect between the kinds of long-term, psychometric studies undertaken in academia and the compliance-oriented priorities that drive regulatory development.

This imbalance points to a possible structural disconnect between academic research and standards development. Rigorous psychometric and perceptual studies are costly in terms of time and are methodologically demanding, yet are essential for evaluating AVAS effectiveness. Such work generally falls outside the remit of industry, where commercial pressures override such pursuits as evidenced in some of the QRTV minutes. This sort of research is more naturally suited to academia, particularly within publicly-funded programs capable of sustaining extended, exploratory investigation. However, while academic studies on EV detectability and sound design have been published, their insights have rarely been integrated into the iterative, real-world processes of standards development – mostly due to a lack of direct academic involvement in the standards process itself.

The reasons for this disconnect are complex. Academic-reward structures often prioritize peer-reviewed publications, citations, and research grants, whereas active participation in standards-bodies is undervalued, despite its strict rigor and real-world impact. Moreover, the timelines and processes of international standards development do not always align neatly with academic cycles of funding and dissemination. The result is a sizeable, missed opportunity - for example, AVAS represents a case where academic evidence could and should be central but instead remains peripheral to the shaping of policy and regulation.

Addressing this gap requires not only greater academic engagement with standards organizations, but also a cultural shift in how research impact is measured. Contributions to standards - which undergo stringent peer review - demand technical precision and directly influence public safety. Therefore, standards contributions should be a recognized and legitimate form of academic output. In the case of AVAS, this alignment is particularly urgent. Without stronger academic participation, the field risks continuing along a path where compliance takes precedence over effectiveness, and where critical safety questions remain unresolved.

E. Limitations of the EVA Study

The EVA study, while providing valuable insights into pedestrian perceptions of AVAS, is subject to several limitations. First, the sample size, although sufficient for statistical analysis, was relatively modest. A larger and more diverse participant pool would strengthen the generalizability of findings, particularly across different demographic groups and geographic contexts. Second, while the survey design captured both quantitative Likert responses and qualitative open-ended insights, future studies could benefit from performance-based experimental designs that combine controlled acoustic testing with lived pedestrian experience. Such triangulation would provide richer evidence for linking perceptual data directly to measurable acoustic and psychometric features.

Another limitation is the scope of documentary analysis of regulatory processes. This paper reviewed selected UNECE QRTV meeting minutes [19] and linked them to subsequent Regulation No. 138, but it did not systematically examine the full corpus of UNECE QRTV and later WP.29 discussions, nor subsequent national transpositions of the regulation. A deeper meta-analysis of these records would allow for a clearer mapping of how specific perceptual concerns were introduced, debated, and either carried through or excluded from the regulatory text.

A further limitation lies in the absence of quantitative data on the degree of academic involvement in standards development. While this paper qualitatively notes that academic input was presented during UNECE's QRTV process but not always operationalized, future work should measure the extent of academic participation across different standards bodies. Such an analysis could also compare how various countries structure, incentivize, or fail to reward academic engagement in standards-making. This would help to clarify whether the observed gaps are structural features of the standards ecosystem itself or instead reflect national priorities and institutional cultures.

These limitations highlight the need for future work that scales up EVA-style studies, expands the review of regulatory archives, and systematically analyses the structural role of academia in standards development. Together, these steps would provide a more comprehensive evidence base for embedding pedestrian-centered perceptual research into future revisions of AVAS regulations.

IX. Conclusions

The EVA study highlights clear and consistent differences in how sighted and vision-impaired pedestrians experience the sounds of EVs. While current AVAS regulations ensure baseline audibility and compliance, they do not adequately address the functional needs of pedestrians who depend primarily on auditory cues. Across both groups, but especially among participants with vision-impairments, experiences of surprise and difficulty in anticipating vehicle actions point to fundamental gaps in the design and evaluation of AVAS.

These findings underscore three interlinked challenges for the future of AVAS: the absence of clear guidance on notification timing and effective distance; the tendency of many current implementations to prioritize branding over functional meaning; and the lack of consideration for the cumulative urban soundscape created by large-scale deployment of synthetic tones. Together, these gaps illustrate why compliance-based testing, while necessary, is insufficient for ensuring real-world safety and accessibility.

Evidence in the historical UNECE QRTV minutes shows that these gaps are not new concepts or phenomena. One particularly striking example came from D'Angelico (QRTV-06-08e, San Diego, 2011) [19], who focused on extrapolating how ICE vehicles naturally convey low-frequency modulations (40 Hz–130 Hz) that pedestrians intuitively associate with starting, stopping, or approaching. This, and other high-quality submissions - such as Pedersen's filter model incorporating perceptual hooks, and the European Association of Automotive Supplier's (CLEPA) focus on locatability - show that perceptual meaning was recognized as an essential dimension during the QRTV process, but was not ultimately operationalized in the regulation. Another example comes from the University of Duisburg-Essen (QRTV-09-02, Bonn, 2011) [19], which undertook a multi-phase study involving 240 participants, including blind and hearing-impaired groups. Their findings identified the 10–20 km/h "danger zone," where BEVs were consistently perceived as riskier than ICE vehicles despite only small physical SPL differences. Blind participants reported strong concerns about inaudibility, late detection, and potential social exclusion, with a clear preference for ICE-like cues over novel EV tones. Crucially, this study highlighted that perceived safety depends as much on predictability and familiarity as on raw sound levels - a conclusion echoed in EVA's survey responses more than a decade later.

It is also worth recognizing that such early contributions in 2010 and 2011 flagged issues such as quiet ICE comparators, e-bike inclusion, and notification distance, yet many of these concerns were only partially reflected in the published Regulation No. 138, which as already mentioned, centered primarily on SPL and frequency thresholds. The EVA results therefore confirm unresolved issues that were identified in UNECE's formative years but not fully carried through into regulation, highlighting the continuing need for pedestrian-centered evidence to guide future revisions.

Equally important is the structural gap between academic research and the standards-making processes that eventually help shape AVAS regulation. To date, regulation has been largely industry-driven, with limited incorporation of empirical insights from up-to-date perceptual and psychometric research. Yet it is precisely this kind of research that is needed to ensure that AVAS functions as an effective communication tool rather than a mere regulatory requirement. Strengthening academic participation in standards would not only bring additional scientific rigor to policy design but would also ensure that all pedestrian groups are properly represented in regulatory and policy frameworks.

The EVA study, in a small way, illustrates the value of bridging these worlds. It demonstrates how collaboration between public bodies, research funders, and academia can connect data collection to lived pedestrian experiences while framing those insights within the context of standards, regulation, and policy. Such partnerships show how publicly funded research can address immediate safety-critical issues and help close the gap between research evidence, policy, and applied implementation.

X. Future Work

Future work on EVA will move in several complementary directions. The first is an expanded meta-analysis of regulatory history, examining the full record of QRTV and WP.29 submissions and minutes, and mapping how these inputs were distilled in the final text of UNECE Regulation No. 138 and subsequent EU legislation. Particular attention will be given to how these influenced the ISO and CEN standards, and how they themselves have fed back into regulatory updates. This historical–regulatory mapping will clarify how early perceptual concerns were translated, diluted, or excluded in the journey from submission to regulation or standard.

Equally, there is a need to develop and validate perceptual metrics that can be operationalized within standards compliance testing. One promising pathway is to revisit earlier proposals raised during the QRTV sessions, such as locatability hooks and distance-based perceptual measures, and explore how these can be translated into measurable standards today. By reaching out to the researchers and organizations that contributed such ideas in those formative years, academia can help refine and extend these approaches into robust, standardized protocols with a decade of additional data now available. In this way, future revisions of AVAS regulation could move beyond compliance with minimal thresholds and begin to reflect more of a pedestrian-centered framework grounded in "perceptual effectiveness".

In addition to performance-based user studies, it will be important for the next phase of EVA to expand its methodological scope. Beyond surveys, it will be important to also conduct interviews and focus-group sessions with pedestrians across various demographics to allow for richer, qualitative engagement with user-perspectives. Braun and Clarke's guidelines [66] will be applied in this context to generate themes grounded in participants' lived-experiences, providing deeper insights into how AVAS is interpreted in real urban environments.

A further direction in EVA will focus on academic engagement in standards development, since many of the problems highlighted in this article may perhaps be a symptom (at least in part) of a disconnect between academic outputs and standards development. A systematic analysis of the extent, type, and impact of academic contributions to AVAS standards will be undertaken, contrasting them with more industry-led inputs. This will include exploring how psychoacoustic and perceptual data can be translated into measurable, compliance-ready metrics, bridging the gap between complex psychometric research and regulatory enforcement. Particular emphasis will be placed on whether universal-design principles, as promoted by public authorities and disability organizations, can provide a framework for embedding the perspectives of end-users and stakeholders more consistently into standards processes.

Looking further ahead, future AVAS design may be enhanced through the integration of artificial intelligence and context-aware technologies. An AI-enabled AVAS could dynamically adapt its signals based on environmental noise, pedestrian density, time of day, or the presence of vulnerable users, ensuring notifications are both perceptually effective and socially considerate. Such systems could form part of broader connected-vehicle and smart city infrastructures, combining acoustic cues with multimodal safety-channels. Rather than static compliance artefacts, AVAS could evolve into adaptive, intelligent communication systems that anticipate and respond to real-world contexts, providing a more inclusive and effective safeguard for pedestrians as urban mobility transforms.

Acknowledgment

This publication has emanated from research conducted with the financial support of Taighde Éireann – Research Ireland under Grant number 23/PSF/12154.

Taighde Éireann
Research Ireland

REFERENCES

[1] F. Neff, D. Fitzpatrick, G. Craddock, C. Cheevers, and A. Conway, “Assessing the Effectiveness of Acoustic Vehicle Alerting Systems (AVAS) for Pedestrians with Visual Disabilities,” in *the Tenth International Conference on Universal Accessibility in the Internet of Things and Smart Environments*, IARIA Congress, Nice, France, 2025.

[2] P. J. Edwards, S. E. Moore, and C. Higgins, “Pedestrian safety on the road to net zero: cross-sectional study of collisions with electric and hybrid-electric cars in Great Britain,” *J Epidemiol Community Health,* vol. 78, no. 8, pp. 487-492, 10 July 2024.

[3] Y. Cai, “Pedestrian auditory perception of approaching vehicles from behind in shared space: The impact of quietness of electric vehicles,” *Transportation research procedia,* vol. 78, pp. 594-601, 2024.

[4] F. Mocholí Belenguer, A. Martínez-Millana, F. S. Castells Ramón, and A. Mocholí-Salcedo, “The Effectiveness of Alert Sounds for Electric Vehicles Based on Pedestrians’ Perception,” *IEEE Transactions on Intelligent Transportation Systems,* vol. 23, no. 4, pp. 2956-2965, April 2022.

[5] J. Krautwurm, D. Oberfeld-Twistel, T. Huisman, M. Mareen Maravich, and E. Altinsoy, “Quality Comparison of Dynamic Auditory Virtual-Reality Simulation Approaches of Approaching Vehicles Regarding Perceptual Behavior and Psychoacoustic Values,” *Acoustics,* vol. 7, no. 1, 2025.

[6] S. S. Dahl, J. Oddershede, J. Brunskog, C. P. Volk, and M. C. Van Eeckhoutte, “Noise differences between electric and conventional cars at low speeds,” in *Proc. DAS|DAGA 2025*, pp. 1483–1486, 2025, doi: 10.71568/dasdaga2025.595.

[7] M. Roan, L. Neurauter, D. Moore, and D. Glaser, “Assessing the Ability of Pedestrians to Safely Hear Approaching Electric Vehicles,” *Sound and Vibration,* vol. 51, no. 10, pp. 6-13, 2017.

[8] M. Pardo-Ferreira, J. Torrecilla-García, C. Heras-Rosas, and J. Rubio-Romero, “New Risk Situations Related to Low Noise from Electric Vehicles: Perception of Workers as Pedestrians and Other Vehicle Drivers,” *International Journal of Environ. Res. Public Health,* vol. 17, no. 6701, 2020.

[9] D. Oberfeld and T. Huisman, “Pedestrians’ street-crossing decisions compared between conventional and electric vehicles,” in *NOISE-CON proceedings*, New Orleans, USA, 2024.

[10] M. Wessels, S. Kröling, and D. Oberfeld, “Audiovisual time-to-collision estimation for accelerating vehicles: The acoustic signature of electric vehicles impairs pedestrians' judgments,” *Transportation Research Part F - Traffic Psychology and Behaviour,* vol. 91, pp. 191-212, 2022.

[11] L. Müller and W. Kropp, “Auralization of electric vehicles for the perceptual evaluation of acoustic vehicle alerting systems,” *Acta Acustica,* vol. 8, no. 27, 2024.

[12] P. R. DeLucia, D. Oberfeld, J. K. Kearney, M. Cloutier, A. M. Jilla, A. Zhou, S. Trejo Corona, J. Cormier, and A. Taylor, “Time-to-Collision Estimation With Age-Related Macular Degeneration Using Visual and Auditory Cues: Which Cues are Most Important?,” *Proceedings of the Human Factors and Ergonomics Society Annual Meeting,* 2025.

[13] M. Roan, M. Neurauter, D. Moore, and D. Glaser, “Electric Vehicle Detectability: a Methods-Based Approach to Assess Artificial Noise Impact on the Ability of Pedestrians to Safely Detect Approaching Electric Vehicles,” *SAE International Journal of Vehicle Dynamics, Stability, and NVH,* vol. 1, no. 2, pp. 352-361, 2017.

[14] T. Berge, “Experience and Perception of AVAS on Electric Vehicles in Norway,” in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings, InterNoise 18*, Chicago, USA, 2018.

[15] Eurostat, “Passenger cars in the EU - Statistics Explained,” European Commission, Brussels, 2024.

[16] European Commission, “(EU) 2017/1576 amendment of Regulation (EU) No 540/2014 of the European Parliament and of the Council as regards the Acoustic Vehicle Alerting System requirements for vehicle EU-type approval,” *Official Journal of the European Union,* vol. L 239, no. Sept., pp. 3–6, 2017.

[17] European Parliament and Council, “Regulation (EU) No. 540/2014 of 16 April 2014 on the sound level of motor vehicles and of replacement silencing systems,” Official Journal of the European Union, Brussels, 2014.

[18] United Nations Economic Commission for Europe (UNECE), “Regulation No. 138: Uniform provisions concerning the approval of quiet road transport vehicles with regard to their reduced audibility,” UNECE, Geneva, Switzerland, 2016.

[19] UNECE, “QRTV Previous Sessions - UNECE Transport Division Wiki,” [Online]. Available: https://wiki.unece.org/display/trans/QRTV+Previous+Sessions. [Accessed 10 09 2025].

[20] U. T. Division, “World Forum for Harmonization of Vehicle Regulations (WP.29),” [Online]. Available: https://unece.org/transport/vehicle-regulations/world-forum-harmonization-vehicle-regulations-wp29. [Accessed 18 August 2025].

[21] U.S. Department of Transportation, National Highway Traffic Safety Administration, “Federal Motor Vehicle Safety Standards; Minimum Sound Requirements for Hybrid and Electric Vehicles. Proposed Rule,” *49 CFR Parts 571 and 585, Docket No. NHTSA-2011-0100, Fed. Reg.,* vol. 78, no. 9, pp. 2797–2817, 2013.

[22] A. Fiebig, “Electric vehicles get alert signals to be heard by pedestrians: Benefits and drawbacks,” *Acoustics Today,* vol. 16, no. 4, pp. 20-28, 2020.

[23] National Highway Traffic Safety Administration, *Federal Motor Vehicle Safety Standard No. 141: Minimum sound requirements for hybrid and electric vehicles Code of Federal Regulations, Title 49, Part 571.141,* Washington, D.C., USA: National Highway Traffic Safety Administration, 2018.

[24] Standardization Administration of China (SAC) and the State Administration for Market Regulation (SAMR), *GB/T 37153-2018 - Acoustic Vehicle Alerting System of Electric Vehicles Running at Low Speed,* Beijing: Standardization Administration of the People's Republic of China, 2018.

[25] International Organization for Standardization, *ISO 16254:2024 - Acoustics - Measurement of sound emitted by road vehicles of category M and N at standstill and low speed operation - Engineering method,* Geneva, Switzerland: ISO, 2024.

[26] A. Hastings, C. Scarpone, B. Samiljan, L. Garay-Vega, and J. K. Pollard, “Acoustic Characteristics of Hybrid Electric Vehicles and Safety of Blind Pedestrians,” *Transportation Research Record,,* vol. 2248, no. 08, pp. 61-67, 2011.

[27] U. Sandberg, L. Goubert, and P. Mioduszewski, “Are vehicles driven in electric mode so quiet that they need acoustic warning signals,” in *Proceedings of 20th International Congress on Acoustics, ICA 2010*, Sydney, Australia, 2010.

[28] A. Fortino, L. Eckstein, J. Viehöfer, and J. Pampel, “Acoustic Vehicle Alerting Systems (AVAS) - Regulations, Realization and Sound Design Challenges,,” *SAE International Journal of Passenger Cars—Mechanical Systems,* vol. 9, no. 3, pp. 995-1003, 2016.

[29] A. Pyzik and P. Larsson, “Acoustic vehicle alerting system and method”. United States of America Patent 11,904,772 B2, 20 Feb 2024.

[30] E. M. Fontecha, G. Waters, and T. R. Stuart, “Performance and acceptability of the TfL Urban Bus Sound (AVAS),” in *INTER-*

NOISE and NOISE-CON Congress and Conference, InterNoise22, Glasgow, Scotland, 2022.

[31] B. Kullukcu, M. Engel, and S. Atamer, "Psychoacoustic Modelling of AVAS Sounds: Consumer-Centric Semantic Attribute Development for Electric Vehicles," in *11th Convention of the European Acoustics Association, Forum Acusticum, EURONOISE*, Málaga, Spain, 2025.

[32] T. Walton, A. J. Torija, and A. S. Elliott, "Development of electric scooter alerting sounds using psychoacoustical metrics," *Applied Acoustics,* vol. 201: 109136, December 2022.

[33] M. A. Pallas, M. Capatina, A. Le Bellec, S. Bianchetti and J. Cesbron, "Environmental noise from alerting signals (AVAS) of electric vehicles: comparison with noise prediction models at urban speeds,," in *10th Convention of the European Acoustics Association Forum Acusticum 2023*, Turin, Italy, 2023.

[34] F. Gutierrez, X. Robin, and A. Charon, "Striking the right note: balancing AVAS in tomorrow's soundscapes," in *NOISE-CON*, New Orleans, USA, 2024.

[35] M. Hsieh, H. Chen, M. Tong, and C. Yan, "Effect of Environmental Noise, Distance and Warning Sound on Pedestrians' Auditory Detectability of Electric Vehicles," *International Journal of Environmental Research and Public Health (IJERPH),* vol. 18, no. 17: 9290, 2 September 2021.

[36] D. Oberfeld, M. Wessels, and T. Huisman, "Auditory and audiovisual time-to-collision estimation and road-crossing decisions," in *10th Convention of the European Acoustics Association*, Turin, Italy, 2023.

[37] R. Sanz-Segura and E. Manchado-Pérez, "Does electric-mobility sound futuristic? conceptualization design strategies for the auditory vehicle alerting system (avas)," *DYNA,* vol. 96, no. 1, pp. 26-29, 2020.

[38] S. Lenzi, J. Sádaba, and P. Ciuccarelli, "A designerly approach to the sonification of electric vehicles," in *Design, Research, Society: DRS2022*, Bilbao, Spain, 2022.

[39] S. Lenzi and G. D'Amico, "Sound for Electric Vehicles," in *Audio Branding Yearbook*, Berlin, Germany, Nomos Verlag, 2010.

[40] A. Zeitler, "Psychoacoustic requirements for warning sounds of quiet cars," *SAE International Journal of Passenger Cars - Electronic and Electrical Systems,* vol. 5, no. 2, pp. 572–578, 2012.

[41] S. K. Lee, S. M. Lee, T. Shin, and M. Han, "Objective evaluation of the sound quality of the warning sound of electric vehicles with a consideration of the masking effect: Annoyance and detectability,," *International Journal of Automotive Technology,* vol. 18, no. 4, pp. 699–705, 2017.

[42] K. Yamauchi, T. Sano, S. Hasegawa, F. Tamura, and Y. Takeda, "Detectability and hearing impression of additional warning sounds for electric or hybrid vehicles," in *INTERNOISE 2014 - 43rd International Congress on Noise Control Engineering: Improving the World Through Noise Control*, Melbourne, Australia, 2014.

[43] N. Misdariis, A. Cera, R. Barthelemy, L. Worms, and L.-F. Pardo, "Collaborative sound design for automotive applications - the sound for electric vehicle as an emblematic use case," in *NOISE-CON*, Nantes, France, 2024.

[44] N. Misdariis and A. Cera, "Knowledge in Sound Design: The Silent Electric Vehicle—A Relevant Case Study," in *Proceedings of the Conference on Design and Semantics of Form and Movement - Sense and Sensitivity*, Eindhoven, Netherlands, 2017.

[45] T. Belschner and M. Bodden, "Comprehensive automotive active sound design-part 2: Operational sounds and brand sound.," in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, Melbourne, Australia, 2014.

[46] M. Carron, F. Dubois, N. Misdariis, C. Talotte, and P. Susini, "Designing sound identity: providing new communication tools for building brands "corporate sound"," in *9th Audio Mostly: A Conference on Interaction With Sound*, New York, USA, 2014.

[47] L. Zattra, N. Misdariis, F. Pecquet, N. Donin, and D. Fierro, "Analysis of Sound Design Practices [ASDP]: Research Methodology," in *XXII Colloquio di Informatica Musicale (Congres d'Informatique Musicale CIM)*, Udine, Italy, 2018.

[48] M. Liljedahl and J. Fagerlönn, "Methods for sound design: a review and implications for research and practice," in *5th Audio Mostly Conference: A Conference on Interaction with Sound*, New York, USA, 2010.

[49] N. Misdariis, A. Gruson, and P. Susini, "Detectability study of warning signals in urban background noises: A first step for designing the sound of electric vehicles," *Journal of the Acoustical Society of America,* vol. 19, no. 1, 2013.

[50] R. Stock, "Sonic e-mobility: traffic noise, sound-producing electric vehicles, and blind pedestrians," *Mobilities,* vol. 20, no. 3, pp. 410–426, 05 Jan 2025.

[51] M. Allman-Ward, R. O. Williams, T. Heinz, and M. Demontis, "Designing and delivering the right sound for quiet vehicles,," in *Inter-noise 2014*, Melbourne, Australia, 2014.

[52] C. Kirsch, A. Baki Kocaballi, A. Johnston, A. Naweed, and I. Stevenson, "Co-Designing the Sound of Safety: Embracing Complexity in the Acoustic Vehicle Alerting System Sound for Zero Emission Buses," *Human Factors and Ergonomics in Manufacturing & Service Industries,* vol. 35, no. 3, 2025.

[53] P. Fink, S. Doore, X. Lin, M. Maring, P. Zhao, A. Nygaard, G. Beals, R. Corey, R. Perry, K. Freund, V. Dimitrov, and N. Giudice, "The Autonomous Vehicle Assistant (AVA): Emerging technology design supporting blind and visually impaired travelers in autonomous transportation," *International Journal of Human-Computer Studies,* vol. 179, 2023.

[54] F. M. Belenguer, A. Martinez-Millana, F. S. C. Ramón, and A. Mocholi-Salcedo, "The effectiveness of alert sounds for electric vehicles based on pedestrians' perception," *IEEE Transactions on Intelligent Transportation Systems,* vol. 23, no. 4, pp. 2956-2965, 2020.

[55] J. F. Petiot, B. G. Kristensen, and A. Maier, "How should an electric vehicle sound? User and expert perception," in *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Portland, Oregon, USA, 2013.

[56] E. Zwicker and H. Fastl, Psychoacoustics: Facts and Models, Berlin, Germany: Springer, 1999.

[57] N. Yasui and M. Miura, "Design of alert sound for electric vehicle based on fluctuation strength for amplitude fluctuation," in *INTER-NOISE and NOISE-CON Congress and Conference, InterNoise23*, Chiba, Japan, 2023.

[58] K. Yamauchi, "Effectiveness of additional warning sounds for quiet vehicles in urban noise environment," *Journal of the Acoustical Society of America,* vol. 140, no. 4, November 2016.

[59] T. Souaille, "An interactive bi-objective optimisation process to guide the design of electric vehicle warning sounds," *Design science,* vol. 8, Jan 2022.

[60] M. Calloni, R. Hadjit, P. Salvekar, and C. Musser, "Harmonizing Safety Regulations and Perception: A Simulation Methodology for AVAS System Design," *SAE technical paper series,* Vols. No. 2024-01-2915, June 2024.

[61] International Organization for Standardization, *ISO 10844:2021 – Acoustics — Specification of test tracks for measuring noise emitted by road vehicles,* Geneva, Switzerland: ISO, 2021.

[62] National Disability Authority, "Advice Paper on Disability Language and Terminology," 2023. [Online]. Available: https://nda.ie/uploads/publications/NDA-Advice- Paper-on-Disability-Language-and-Terminology.docx. [Accessed April 2025].

[63] H. McHugh and F. Dolan, "Guidelines for Good Adult Literacy Work," National Adult Literacy Agency, Dublin, Ireland, 2012.

[64] F. Neff, D. Fitzpatrick, G. Craddock, and C. Cheevers, "Making Electric Vehicles Audible: Perception and Inclusivity in Acoustic Vehicle Alerting System Design," in *30th Association for the Advancement of Assistive Technology in Europe (AAATE)*, Nicosia, Cyprus, 2025.

[65] L. Müller, J. Forssén, and W. Kropp, "Auditory localization of multiple stationary electric vehicles," *The Journal of the Acoustical Society of America,* vol. 157, no. 3, pp. 2029–2041, 2025.

[66] V. Braun and V. Clarke, "Using thematic analysis in psychology.," *Qualitative Research in Psychology,* vol. 3, no. 2, pp. 77–101, 2006.