Advances in Science and Technology Research Journal

Advances in Science and Technology Research Journal, 2027, 21(1), 258–272
https://doi.org/10.12913/22998624/231694
ISSN 2299-8624, License CC-BY 4.0

Received: 2026.07.23
Accepted: 2026.11.15
Published: 2026.12.01

Reaction time to hazardous acoustic signals as a design benchmark for a local assistive system for people with hearing loss

Jakub Krzysiak[1*], Andjelka Hedrih[2], Rafał Rusinek[3], Michał Wydra[1]

[1] Department of Applied Computer Science, Faculty of Mathematics and Information Technology, Lublin University of Technology, Nadbystrzycka 38, 20-618 Lublin, Poland
[2] Department of Mechanics, Mathematical Institute of Serbian Academy of Sciences and Arts, Kneza Mihaila 36, 11000 Belgrade, Serbia
[3] Department of Applied Mechanics, Faculty of Mechanical Engineering, Lublin University of Technology, Nadbystrzycka 36, 20-618 Lublin, Poland
* Corresponding author's e-mail: j.krzysiak@pollub.pl

ABSTRACT

The ongoing development of acoustic event recognition methods and Sound Event Detection (SED) systems creates new opportunities to support the safety of people with hearing loss. This paper establishes a behavioral reference point for the design of such systems by measuring, in a controlled field experiment, how people respond to hazardous acoustic signals without assistance. Sixty participants took part, in two groups of thirty, one reporting no hearing loss and one with hearing loss, and each completed 20 stimulus exposures, giving 1200 trials in total. Two outcomes were analyzed separately: whether a response was recorded at all, and, conditional on a response, how long it took. In the hearing-loss group, 49.2% of trials produced no recorded response, compared with 0.0% in the comparison group, a difference of 49.2 percentage points (95% confidence interval [46.0, 52.3]) amounting to complete separation between the groups, as every participant with hearing loss failed to respond to at least 30% of signals while no comparison participant missed any. Evaluated against the standardized 5500 ms observation window, the no-response rate was 52.3%. Among trials with a recorded response, the median reaction time was 502 ms in the comparison group and 670 ms in the hearing-loss group. These results indicate that the principal design challenge is not only the speed of the response but the frequency with which a hazardous signal produces no behavioral response at all. On this basis the paper formulates provisional design hypotheses for the AI EAR SUPPORT system, covering on-device inference and multi-channel notification through a smartphone and a smartwatch. No artificial intelligence model, notification interface, or assisted user response was tested in this study, and the hypotheses therefore require empirical validation.

Keywords: artificial intelligence, assistive systems, hazardous sounds, sound event detection, reaction time, hearing loss, on-device inference, acoustic safety.

INTRODUCTION

Sound is one of the basic sources of information about the environment. In hazardous situations, it can serve as an early warning signal, informing about fire, an approaching emergency vehicle, a scream, a car horn, a technical alarm, or other events requiring an immediate response. For people with normal hearing, this information is received almost automatically, and auditory cues have been shown to accelerate detection and judgment compared with vision alone [1]. For people with hearing loss or other hearing disorders, access to this layer of environmental information is limited, which can reduce safety at home, in public spaces, and in workplaces.

Assistive listening technologies as a category have historically suffered from limited real-world adoption, often due to usability, trust, and design barriers rather than a lack of underlying technical capability [2, 3].

In recent years, there has been strong development in acoustic event recognition methods. Sound Event Detection systems not only classify the type of event, but also aim to determine its onset and offset in the signal. Approaches based on domain generalization frameworks have demonstrated that SED models can be made robust to real-world acoustic conditions by integrating heterogeneous training data. Yet, performance across unseen environments remains a challenge [4]. Large-scale pre-trained neural networks, trained on synthetic datasets of over 1000 hours of labeled audio, have substantially improved SED generalization capabilities [5]. Meanwhile, semi-supervised methods exploiting both labeled and unlabeled data have achieved state-of-the-art performance on benchmark datasets such as DESED [6].

At the same time, the development of this field has been oriented primarily towards improving classification and detection performance, rather than linking model results to the end user's biological reaction time. Even recent domain-generalization approaches – for example, mean-teacher frameworks trained on heterogeneous data [4] – primarily target improved recall on unseen event categories rather than end-to-end response latency. Simple and choice reaction-time physiology is well characterized in normal-hearing populations [7]. Comparable benchmarks also exist for related human-machine safety contexts, most notably the takeover-time literature for automated vehicles [8, 9]. Separately, a growing body of work documents compensatory enhancements in visual attention and reactivity among deaf and hard-of-hearing individuals [10, 11]. Related evidence indicates that visual motion is represented in auditory and association cortices in early-deaf individuals [12]. Neither line of research establishes how quickly – or whether at all – a person with hearing loss detects and reacts to a hazardous acoustic event without assistance; a recent systematic review of AI-based applications for the deaf and hard-of-hearing population did not identify real-time environmental hazard alerting among the AI applications it included [13]; that review applied a narrow search strategy requiring "AI" or "artificial intelligence" in the title, so it cannot establish the absence of such work. The paper aims to propose a benchmark of reaction time to hazardous acoustic events and to indicate how this benchmark can be used as a design parameter for the AI EAR SUPPORT system. This methodological and applied study estimates reference human response times and examines how they can inform the design of low-latency, on-device assistive systems.

Existing research on SED primarily focuses on model architectures, acoustic feature representations, data augmentation, and segmentation and classification metrics. The literature describes both older approaches based on Hidden Markov Models/Gaussian Mixture Models (HMM/GMM) and Non-negative Matrix Factorization (NMF), and modern deep models including Convolutional Recurrent Neural Networks (CRNN) and large Convolutional Neural Networks (CNN) pre-trained on AudioSet [14, 15]. Synthetic datasets such as DESED and URBAN-SED [16] are also being developed, obtained by combining acoustic backgrounds with anomalies and precise onset/offset labeling. Noise-robust SED remains a particularly active research direction, with recent work combining large language models with audio source separation to handle overlapping events [17]. Lightweight architectures employing depth-wise separable convolutions have demonstrated that compact CRNN models – using only 3 % of standard parameter counts – can achieve competitive performance on domestic safety event detection tasks [18].

However, it is still difficult to find studies that simultaneously:

- empirically determine human reaction time to dangerous acoustic stimuli,
- compare individuals without reported hearing loss and individuals with hearing disorders,
- treat the obtained reaction times as a design benchmark for AI systems,
- integrate event detection with a practical notification layer via mobile devices and smartwatches.

This gap is significant from the perspective of designing assistive systems. A highly accurate detection model may prove insufficient if the time to inference, signal aggregation, and alert generation is too long relative to the user's needs, especially when installed on mobile devices. Detection accuracy and processing speed are not independent design parameters: architectures that achieve higher classification accuracy typically

rely on greater model depth or more complex feature aggregation, both of which increase computational cost and inference latency [19, 20]. Without an explicit latency budget grounded in human reaction time, model selection tends to optimize accuracy in isolation, risking systems that are highly accurate but too slow to deliver a functional safety benefit. The research gap concerns the insufficient linkage between the technical metrics of acoustic event detection and the system's functional effect, understood as a reduction in user response time in a potentially dangerous situation.

Performance in noisy environments remains a major unresolved challenge for SED systems. Noise-robust fine-tuning using large language model-assisted noise augmentation [17] and semi-supervised consistency regularization methods [6] have improved results on noisy benchmark splits. Still, the gap between laboratory and real-world performance persists. Curriculum learning approaches that order training examples by detection difficulty have improved F-scores by over 10 percentage points on standard datasets [21], yet do not address the fundamental problem of missing annotations for safety-critical events. Earlier work on alarm and siren detection specifically for deaf and hard-of-hearing users has demonstrated technical feasibility, including real-time embedded and on-device smartphone implementations [22, 23], but without grounding detection or alert-timing requirements in empirically measured human reaction times. Among the systems reviewed here, we found limited evidence addressing the human interface layer more broadly, that is, how quickly information about a hazard should be conveyed to the user and how a detection result should be converted into an alert that a person with hearing loss can act upon.

The development of SED progressed from HMM/GMM and NMF methods to deep CRNN and large CNN networks pretrained on AudioSet [5, 6]. Pre-trained networks built on large-scale synthetic datasets – encompassing 1167 hours of audio across 170 sound classes – have demonstrated strong generalization and are now considered a foundation-model approach in the field [5]. Datasets for SED are increasingly synthetic: environmental backgrounds are combined with anomalies from labeled monophonic databases to generate samples with precise onset/offset labels. For the AI EAR SUPPORT project, a focused set of application-oriented classes was adopted: background sounds, alarms, dog barking, car horns, screaming and shouting, and sirens. This selection is consistent with the safety event taxonomies reported in the literature, which typically include sirens, screams, alarms, and car horns as primary hazard categories [4, 18].

Compact model architectures deserve special attention in the context of on-device deployment. Low-complexity CRNN models based on depthwise separable convolutions have been shown to achieve performance comparable to full-parameter systems while using only a fraction of the computational resources [18]. This is directly relevant to the AI EAR SUPPORT target deployment scenario on mobile devices and smartwatches.

MATERIALS AND METHODS

The measurements were conducted in two sessions: one for the group that reported no hearing loss and one for the group with hearing loss. Both sessions used the same purpose-built portable measurement station in an identical configuration, so that stimulus presentation, response capture, and timing were reproduced. The study included N = 60 participants organized into two groups of 30:

- Comparison group (Group A): 30 volunteers without reported hearing loss. Hearing status in this group was self-reported by participants; no audiometric testing was performed.
- Hearing-loss group (Group B): 30 individuals with hearing loss. Because no audiometric data were collected, the degree and type of impairment within this group have not been documented.

Each participant completed 20 stimulus exposures, yielding a total of 1200 trials across the full sample. As noted above, demographic and audiological characteristics – including age, sex, dominant hand, method of audiological qualification, degree and type of hearing loss, and hearing-aid use during testing – were deliberately not collected and therefore could not be included as covariates. This is a constraint inherent to the design-oriented, non-medical scope of the study rather than an omission at the analysis stage. Age in particular would be a relevant covariate, as reaction time has been shown to vary systematically between younger and older adults [24]; the consequences of its absence are addressed in the Limitations section.

Experimental environment

The experiment was designed as a controlled field study, with the setup kept constant across participants to ensure comparable stimulus exposure. Because the measurements were conducted at the participating organizations rather than in a dedicated acoustics laboratory, a portable research station was constructed specifically for this study, enabling an identical measurement setup to be replicated at both sites. Participants were seated in a quiet room facing a screen. Audio stimuli were delivered via a loudspeaker system, consistent with loudspeaker-based protocols used in comparable auditory reaction time studies [25]. Stimuli were reproduced through an active, mains-powered full-range loudspeaker (Behringer Eurolive B205D; 133 mm full-range neodymium driver, integrated 125 W RMS Class-D amplifier, rated frequency response 95 Hz–12 kHz at −3 dB and 65 Hz–18 kHz at −10 dB, maximum sound pressure level 113 dB at 1 m), positioned in front of the seated participant at approximately ear height and on the listening axis, and driven from the line output of the control computer. The playback level was set to 75 dB SPL at the listening position and was held constant across participants, sessions, and sound classes. In addition, all stimulus samples were level-normalized across classes before the experiment, so that differences in reaction time would reflect the perceptual characteristics of the event type rather than differences in presentation loudness; this nominal level served as the normalization reference for the present stage of the work. The frequency weighting, the sound level meter used, and the exact loudspeaker-to-participant distance were not documented, so the reported level should be interpreted as a nominal presentation setting rather than as a calibrated acoustic measurement. The implications of this constraint are addressed in the Limitations section.

Stimulus presentation, trial sequencing, and randomization were controlled by dedicated software running on a computer; the station drew stimulus samples in a randomized order for each experimental run, and all trial events were automatically logged by the system. Participant responses were captured using a dedicated large-dome response button connected via a USB interface, following the general approach used in automated auditory reaction-time measurement systems [26]. Reaction time was operationally defined as the interval between stimulus onset and button press, recorded with millisecond resolution. Trials without a registered response within the observation window were logged as a separate "no response" category rather than being excluded from analysis, allowing the miss rate itself to be quantified as an outcome measure alongside reaction time. Reaction time was measured by the system from the onset of the acoustic stimulus, and the default observation window was 5500 ms from that onset. In individual cases, the coordinator requested that this window be extended, and 19 responses in the hearing-loss group were consequently recorded later than 5500 ms, the longest being 23 seconds. Because the exact closing time of each trial was not logged, 5500 ms is treated throughout as the interval for which observation is complete in every trial, and results are reported against that reference; the sensitivity of the conclusions to the treatment of the 19 later responses is examined in the Results section. This design was deliberately adopted for the first study, whose purpose was to obtain an initial estimate of response times; a more tightly controlled and more fully instrumented protocol is planned for the subsequent studies in this cycle.

Stimulus materials

The stimulus material is catalogued in six classes: background sounds, which serve as a control class; alarms; dog barking; car horns, shown as "honks" in the figures; screaming and shouting; and sirens. Within the collection, each class is stored under a fixed numerical identifier that is used for file organization only and carries no experimental meaning. The present experiment used the five hazard classes – alarms, dog barking, car horns, screaming and shouting, and sirens – each presented 240 times across the full sample, yielding 1200 trials in total. Each stimulus was presented from its onset phase – the key element for measuring reaction to the early phase of the hazardous sound. The response window was measured from stimulus onset, and responses were recorded within the first 5500 ms of playback. Recordings were drawn from the project's Acoustic Training Resources (AZT) collection, assembled from commercial and openly available sound libraries (including Envato Elements [27]), and supplemented by a set of recordings made by the project team for model-training purposes. Stimuli were drawn from a pool of 1055 distinct

audio clips (alarms: 207; dog barking: 203; car horns: 210; screaming and shouting: 217; sirens: 218). Each participant received exactly four presentations per class, that is, 20 trials in total, with clips drawn at random from the pool for each run; consequently, most clips were presented to only one participant. The class taxonomy is consistent with safety-oriented SED taxonomies reported in the literature [4, 18].

RESULTS

Because each participant contributed 20 trials, individual trials are not independent observations; all inferential comparisons were therefore conducted with the participant as the unit of analysis (n = 30 per group). Two outcomes were analyzed separately, reflecting the two-part structure of the response process. First, the no-response rate was computed for each participant and compared between groups using the Mann–Whitney U test, with the rank-biserial correlation as the effect size and a participant-level bootstrap (20 000 resamples) to estimate the confidence interval for the risk difference. Second, reaction time among trials with a registered response was summarized as a per-participant median and compared using the Mann–Whitney U test, with the rank-biserial correlation and the Hodges–Lehmann shift estimator with a bootstrap confidence interval.

As a complementary trial-level model that retains all successful responses, reaction times were log-transformed and analyzed using generalized estimating equations with an exchangeable working correlation structure and participants as the clustering unit, yielding a cluster-robust ratio of geometric means. This model was cross-checked against a linear mixed model with a participant random intercept and against a cluster bootstrap resampling of participants; all three approaches produced concordant estimates. Differences in miss rate across sound classes within the hearing-loss group were tested with the Friedman test, using per-participant, per-class miss rates. The significance level was set at $\alpha = 0.05$.

One consequence of the data structure required a specific analytical decision. No participant in the control group had a single missed trial (0 of 600), indicating complete separation with respect to group membership. An odds ratio for the miss rate is therefore not estimable, as it diverges to infinity. The between-group difference in miss rate is accordingly reported as a risk difference with a confidence interval, rather than as an odds ratio. Any zero-event bound is reported at the participant level, since trials are clustered within participants. Reaction time distributions were characterized using two complementary approaches. In the first, no-response trials were treated as missing data and excluded, allowing the shape and kinetics of successful reactions to be examined through histograms and kernel density estimates (Figure 1).

Figure 1 presents the reaction time distributions for successful responses only. The comparison group (blue, N = 30) shows a compact, unimodal distribution peaking near 400–500 ms. The hearing-loss group (orange, N = 30, successful responses only) shows a substantially flattened and shifted distribution, with a long right-skewed tail extending well beyond 3000 ms, including a small subset of extreme outliers above 6000 ms.

In the second, the same successful responses were summarized as an empirical cumulative distribution function (ECDF), which shows the proportion of recorded responses that had occurred by each point in time and therefore compares the kinetics of the two groups independently of how often a response occurred at all (Figure 2).

In the control group (Group A, N = 30), all 600 trials produced a registered response. The distribution of reaction times was compact and unimodal, characteristic of a standard simple auditory reaction time profile. The key descriptors were a median reaction time of 502 ms and a maximum of 1983 ms. The majority of responses were clustered in the 300–700 ms range. This lies above classic simple auditory reaction time benchmarks, typically reported around 150–300 ms in controlled detection tasks [28]; the difference likely reflects the added stimulus-recognition demand of the present task, in which participants had to identify a sound as hazardous rather than merely detect its onset. This range, anchored by the median of 502 ms, constitutes the primary biological reference point for AI EAR SUPPORT system design.

In the hearing-loss group (Group B, N = 30), the results revealed a qualitatively different profile with direct operational consequences. For successful responses, the median was 670 ms and the mean 1555 ms. The no-response rate reached 49.2% of all 600 trials – meaning that nearly one in two hazardous acoustic signals produced no recorded reaction. This is the most operationally

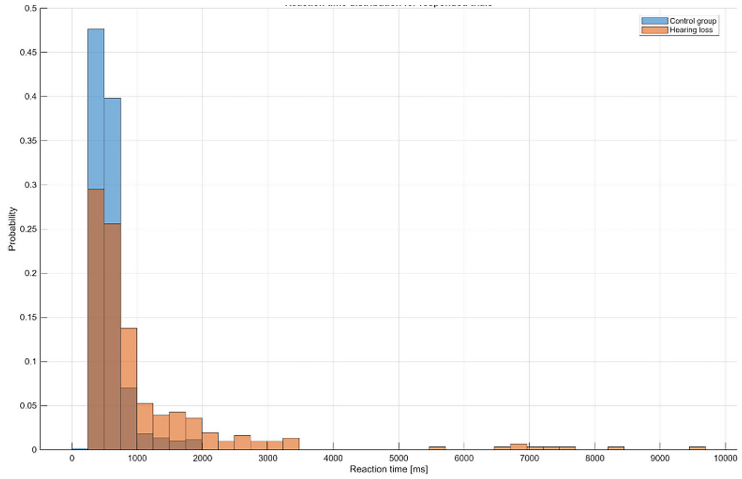


Figure 1. Distribution of recorded reaction times: 1) median 502 ms for the control group; 2) 670 ms for the hearing-loss group

significant finding of the study. The density curve for the hearing-loss group shows a long right-skewed tail extending well beyond 3000 ms, including a small subset of extreme outliers above 6000 ms (Figure 1), consistent with a pattern in which some responses reflect substantially delayed perceptual processing rather than a single characteristic latency.

Figure 2 presents, for each group, the cumulative proportion of recorded responses as a function of time since stimulus onset. The denominator in each curve is the number of trials in which a response was registered, so the figure characterizes response kinetics conditional on a response having occurred and does not itself display the no-response rate; the curves are therefore not survival estimates and involve no censoring. The comparison between them shows that responses in the hearing-loss group accumulated more slowly throughout the observation window. The proportion of trials in which no response occurred is reported separately. Counted against all 600 trials in each group, and evaluated at the 5500 ms default observation window within which observation is complete for every trial, the comparison group had reached 100%, while the hearing-loss group had reached 47.7%; equivalently, 52.3% of hearing-loss trials had no recorded response within the default window. This standardized figure is slightly higher than the 49.2% overall no-response rate because 19 responses in the hearing-loss group were recorded later, in trials with extended windows. The overall rate of 49.2% is therefore a conservative, lower-bound

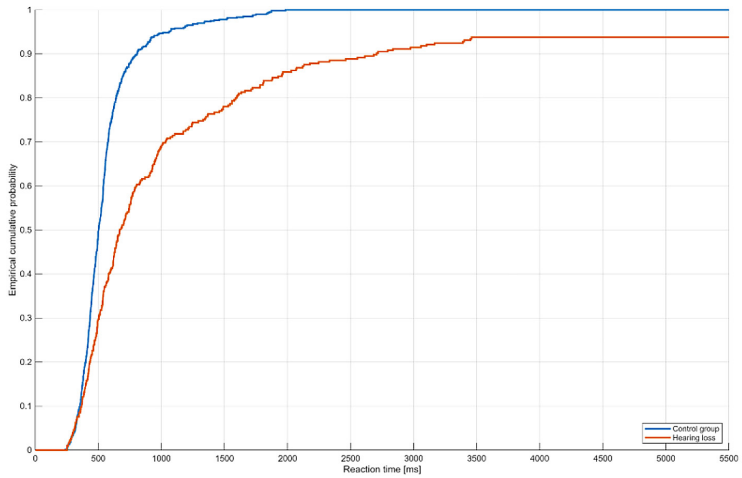


Figure 2. Empirical cumulative distribution of reaction times among trials with a recorded response, for the comparison group and the hearing-loss group. Trials without a recorded response are not represented in this figure; they are reported separately in the text and in Figure 3

estimate of the failure-to-respond proportion. The no-response rate was additionally analyzed as a function of sound class to identify whether specific stimulus types were disproportionately missed (Figure 3).

The comparison between groups demonstrates a clear, practically significant asymmetry. Table 1 summarizes the key descriptive statistics for both groups.

Group comparison

Analyses conducted at the participant level confirm that both differences observed descriptively are statistically robust and that they differ markedly in character. The difference in miss rate is categorical rather than gradual. Evaluated at the default 5500 ms window, within which observation is complete in every trial, 52.3% of hearing-loss trials had no recorded response, against 0% of comparison-group trials; the overall rate of 49.2% reported above is correspondingly a lower bound. Every participant in the hearing-loss group failed to respond to at least 30 % of the signals presented, whereas no participant in the comparison group missed a single signal. The two distributions therefore do not overlap at all on this measure, which is reflected in a rank-biserial correlation of exactly 1.00 – complete separation. The risk difference is 49.2 percentage points (95% CI [46.0, 52.3]). With zero events in the comparison group, the corresponding odds ratio is not estimable. A trial-level rule-of-three bound is not appropriate here because the 600 control trials are clustered within 30 participants rather than being independent. At the participant level, no participant in the comparison group missed any signal (0 of 30), which corresponds to a 95% upper bound of approximately 10% for the proportion of such participants (11.6% by the exact Clopper–Pearson interval).

The difference in reaction time among successful responses is, by contrast, substantial but overlapping. The medians reported in this section are computed at the participant level, that is, as the median of the 30 individual median reaction times in each group, and therefore differ slightly from the pooled trial-level medians of 502 ms and 670 ms reported above: participant-level summaries weight every participant equally, whereas pooled medians are weighted towards participants who contributed more successful responses. Per-participant median reaction times differ by 164.9 ms (Hodges–Lehmann shift, 95% CI [102.9, 235.6]; rank-biserial $r = 0.82$), and the trial-level cluster-robust model estimates that successful responses in the hearing-loss group take 1.66 times longer on the geometric-mean scale (95% CI [1.51, 1.82]). The between-participant variance component in log reaction time was estimated at essentially zero, indicating that trial-to-trial variability within participants dominates individual differences between them, which is consistent with

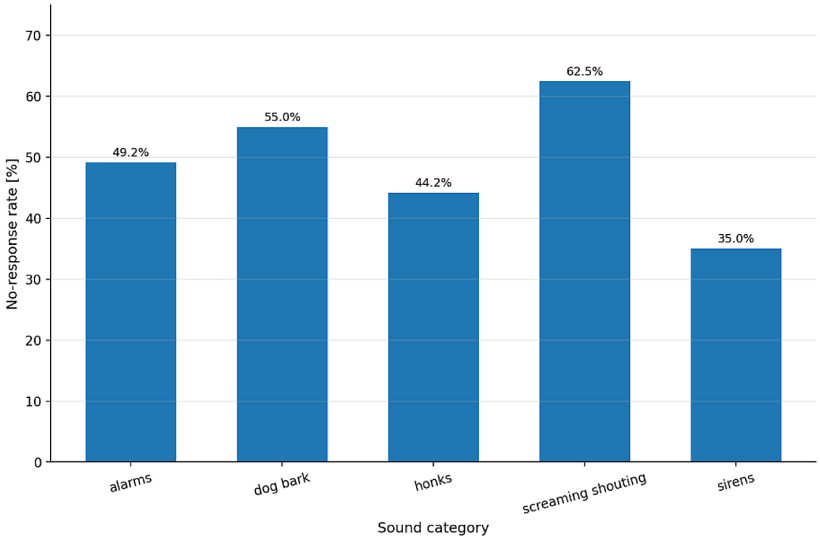


Figure 3. No-response rate by sound class for the hearing-loss group
Screaming/shouting and dog barking show the highest miss rates; sirens show the lowest, despite not producing the fastest successful reaction times

Table 1. Summary of reaction time results by group

Parameter	Group A – Comparison (N = 30)	Group B – Hearing loss (N = 30)	Note
Total trials	600	600	20 per participant
No-response rate	0.0 %	49.2 %	Key safety finding
Median RT (successful)	502 ms	670 ms	
Mean RT (successful)	560 ms	1555 ms	
Maximum RT	1983 ms	no response	Stopped by the operator after 23.6 s
Distribution shape	Compact, unimodal	Wide, right-skewed with a long tail	

Note: RT = reaction time. Maximum RT values include extreme outliers from the hearing-loss group.

the close agreement between the participant-level and trial-level estimates. Table 2 summarizes the comparison.

Because the observation window was extended in a small number of trials, and because the closing time of each trial was not logged, the robustness of these results was checked against three alternative treatments of the 19 hearing-loss responses recorded after the 5500 ms default window. When these responses are retained as recorded, the no-response rate is 49.2%, and the per-participant median reaction time is 674.5 ms. When they are instead reclassified as failures to respond within the default window, the rate is 52.3%, and the median is 609.8 ms. When they are excluded from the analysis altogether, the corresponding values are 50.8% and 609.8 ms. In all three treatments, the separation between groups on the miss rate remains complete (rank-biserial $r = 1.00$, $p < 0.001$) and the difference in reaction time remains significant ($p < 0.001$ in every case). The affected trials represent 1.6% of the full dataset and 3.2% of the trials in the hearing-loss group, that is, in the only group to which they apply. The direction and significance of the between-group conclusions are therefore unchanged across all three treatments, although the numerical estimates, and in particular the median reaction time, do depend on which treatment is adopted.

Within the hearing-loss group, the miss rate also varied systematically across sound classes (Friedman $\chi^2 = 16.16$, $df = 4$, $p = 0.003$; Kendall's $W = 0.135$, indicating a small omnibus effect). This omnibus test does not identify which individual classes differ; the class rankings reported below are therefore descriptive, as no corrected pairwise comparisons were performed. The omnibus result, however, confirms that the heterogeneity across classes is not attributable to chance.

Dependence on sound type

Reaction times varied across sound categories in both groups, which is consistent with evidence that alarm type and acoustic salience systematically affect reaction speed in safety-critical listening tasks [29]. The ordering of categories, however, does not follow perceptual salience in any simple way: in the hearing-loss group the fastest median response among successful trials was obtained for dog barking (565 ms) and the slowest for screaming and shouting (909 ms), even though the latter is the more abrupt and more salient signal. Because the present task measured only the interval from stimulus onset to button press, which combines audibility, attention, recognition, decision, and motor execution, the underlying mechanism cannot be identified from these data. Figure 3 breaks down the no-response rate by sound class for the hearing-loss group. The pattern is heterogeneous: screaming/shouting shows the highest miss rate (62.5%), followed by dog barking (55%), while sirens show the lowest (35%) despite not being the fastest class in terms of successful reaction time. This pattern indicates that the no-response rate and the conditional response time carry partly distinct information: a category that is frequently missed is not necessarily slow to respond to when it is detected. The present protocol does not identify the perceptual or decisional mechanism responsible for these differences, and the possible interpretations are considered in the Discussion. Sirens are the most reliably detected category even when the subsequent response is not the fastest – an observation with direct design implications for which classes most urgently require AI-assisted detection versus AI-assisted response acceleration.

In Figure 4, the corresponding reaction time distributions by sound category for both groups are shown on a logarithmic scale. In the

Table 2. Formal between-group comparison at the participant level (n = 30 per group, 20 trials each)

Outcome	Comparison group (n = 30)	Hearing loss (n = 30)	Test and effect size
No-response rate	0.0 % (0/600 trials); No participant missed any signal	49.2 % (295/600); per-participant range 30.0–70.0 %	Risk difference 49.2 pp, 95 % CI [46.0, 52.3]; Mann–Whitney U = 900, p < 0.001; rank-biserial r = 1.00
Median reaction time (successful responses, per participant)	514.6 ms	674.5 ms	Hodges–Lehmann shift 164.9 ms, 95 % CI [102.9, 235.6]; Mann–Whitney U = 818, p < 0.001; rank-biserial r = 0.82
Reaction time, trial-level model (log scale)	reference	1.66× longer	Ratio of geometric means 1.66, 95 % CI [1.51, 1.82], p < 0.001 (GEE, cluster-robust)

hearing-loss group, median reaction times range from 565 to 909 ms across categories, with screaming/shouting showing both the highest miss rate and the slowest median response among successful trials (909 ms). At the same time, dog barking and car horns show comparatively faster successful responses despite substantial miss rates. This indicates that miss rate and response speed are partially dissociated across categories: a class can be frequently missed without necessarily being slow to respond when it is detected.

DISCUSSION

The results distinguish two complementary outcomes: whether a behavioral response was recorded and, conditional on a response, how long it took. The provisional design hypotheses that follow from these outcomes are collected in Table 3 at the end of this section. At the standardized 5500 ms reference window, 52.3% of trials in the hearing-loss group had no recorded response, compared with 0% in the comparison group; when later responses from extended trials were retained, the overall no-response rate was 49.2%. Participant-level analyses showed complete separation in the no-response rate (rank-biserial r = 1.00) and a large but overlapping difference in successful-response times (Hodges-Lehmann shift of 164.9 ms; 95% CI [102.9, 235.6]). These findings indicate that an assistive system should address both missed events and delayed responses. However, the button-press task combines audibility, attention, recognition, decision, and motor execution. A trial without a recorded response therefore cannot be taken as evidence that the sound was not perceived.

Human reaction time can inform assistive-system design, but it does not by itself define a validated latency requirement. The median values of 502 ms in the comparison group and 670 ms among successful responses in the hearing-loss group are biological reference values rather than direct inference-time thresholds. Evaluation must cover the complete event-to-action chain: acoustic capture, detection, decision, alert delivery, and user response. A functional advantage can be established only by comparing assisted and unassisted performance while jointly measuring end-to-end latency, detection sensitivity, missed detections, false alarms, and alert usability.

The sound-class analysis suggests that miss rate and response speed are partly dissociated. Some classes were frequently missed but elicited relatively fast responses when detected, whereas screaming/shouting combined the highest miss rate with the slowest median successful response. Because the omnibus sound-class effect was small (Kendall's W = 0.135) and no corrected pairwise comparisons were performed, these class-specific patterns should be treated as exploratory. They nevertheless support future testing of class-specific detection thresholds and notification strategies, rather than assuming that a single latency target is equally appropriate for all hazardous sounds.

The contribution of this study should be understood as an integration of research layers rather than as a claim that reaction-time benchmarks or assistive alerting technologies are absent from the literature. Established work provides simple-reaction references [7] and takeover-time references [8, 9]. Earlier systems demonstrate the technical feasibility of alarm and siren detection [22, 23]. The review by Albash [13] did not identify real-time environmental

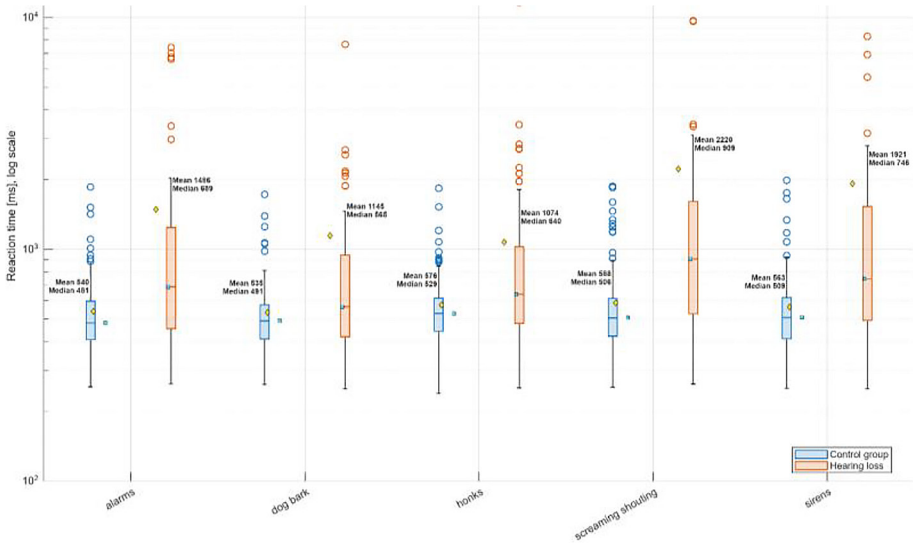


Figure 4. Reaction time by sound category and group, log scale. Boxes show interquartile range; diamonds mark group means, squares mark group medians per category; circles are individual trial outliers

hazard alerting among the applications included. However, its title-based search strategy was too narrow to establish that no such work exists. The present contribution is therefore the empirical linkage of behavioral responses to hazardous acoustic events with an end-to-end evaluation framework that includes both on-device detection and a user-notification layer.

On-device architecture and model selection

YAMNet [30] and PANNs [31] are candidate pretrained backbones for future comparison and transfer learning; neither constitutes a validated AI EAR SUPPORT model at this stage. The human-response benchmark can help allocate an end-to-end latency budget across capture, preprocessing, inference, decision, and alert generation. Still, it should not be used as an isolated inference-time requirement. Model selection should therefore compare accuracy, latency, robustness to noise, memory use, and energy consumption on the target hardware and sound taxonomy. Compact architectures and knowledge-distillation methods may be useful in this trade-off, but their performance must be measured rather than inferred from the human baseline. Lightweight audio classification networks developed specifically for resource-constrained hardware provide a concrete reference point for this comparison, having been shown to retain competitive accuracy at a small fraction of the parameter count of full-scale models [32].

On-device processing remains a reasonable architectural hypothesis because it may reduce cloud-related delay and preserve operation when connectivity is unavailable [33, 34]. It also keeps raw ambient audio on the user's device [35]. However, the power budget, sustained always-on performance, privacy safeguards, classification accuracy, and event-onset-to-alert latency of the implemented system have not yet been established experimentally.

Notification layer

The notification layer is an essential part of the proposed safety chain because a technically correct detection has no functional value unless the alert is perceived and understood. Visual and vibrotactile channels are appropriate candidates for evaluation: studies of deaf populations report advantages in certain forms of visual orienting and alerting [10, 11]. Cortical reorganization of visual motion processing has also been reported in early-deaf individuals [12]. Wearable haptics research indicates that location-coded vibration may be detected more reliably than pattern-coded cues [36]. These findings do not demonstrate enhanced visual or vibrotactile processing in the present sample, because the cited evidence is heterogeneous and often specific to early deafness, sign-language background, or particular attention tasks. Any such evaluation must also account for the practical constraints of vibrotactile testing outside laboratory conditions, including actuator placement, skin coupling, and

the ambient vibration present in real environments, all of which affect detection thresholds and have been documented as recurring sources of error in wearable haptics research [37]. Future assisted-versus-unassisted experiments should therefore compare visual, haptic, and combined notifications, including alert urgency, comprehension, false-alarm tolerance, and performance during divided attention and movement.

Limitations of the proposed method

The study was conducted under controlled field conditions, and a simple button press does not reproduce protective behavior in an actual emergency. Participants with hearing loss were recruited from a single organization, and age, degree, type of hearing loss, audiological qualification, and hearing-aid use were not collected. The group assignment was based on employer information rather than on independent audiometric verification. These factors limit generalizability and prevent analysis of potentially important covariates.

The observation window was not uniform in all trials. Although 5500 ms was the default interval for which observation was complete, 19 hearing-loss responses were recorded during coordinator-requested extensions, and the exact closing time of each trial was not logged. The sensitivity analysis shows that the principal between-group findings are robust to alternative treatment of these trials. Still, the 49.2% overall no-response rate should be interpreted as a conservative lower-bound estimate relative to the standardized 5500 ms rate of 52.3%. In addition, the normalization metric and tolerance, measurement instrumentation, frequency weighting, exact loudspeaker distance, and complete clip-duration information were not documented. The end-to-end timing of the measurement station was likewise not validated: the audio-output buffer delay, the polling interval of the USB response button, and the interval between the software playback command and the physical onset of sound in the room were not measured. Reaction times are therefore recorded with millisecond resolution but carry an unquantified constant system offset. Because both groups were measured on the same station in the same configuration, this offset does not affect the between-group comparison, but it does constrain the interpretation of the absolute values as a reference point. The reported playback level should consequently be treated as approximate, and subsequent studies should use a fixed, fully logged observation window and a reproducible acoustic-calibration protocol.

Participation was voluntary, and instructions were provided in Polish Sign Language; however, consent was documented at the group level rather than the individual level, and no individually signed informed consent forms were collected. This limits the formal documentation of voluntariness and should be addressed in future studies through individual informed consent procedures and a documented institutional ethics determination. The present results also do not validate the AI EAR SUPPORT system. The design hypotheses in Table 3 require prospective testing of classification accuracy, inference, and end-to-end latency, false alarms, missed detections, notification effectiveness, and assisted user performance.

Table 3. Provisional design hypotheses for AI EAR SUPPORT, derived from the present behavioral benchmark and requiring empirical validation

KPI	Baseline – without system	Provisional design hypothesis (requires validation)
No-response (miss) rate	49.2% (hearing-loss group)	Hypothesis: substantial reduction of the standardized no-response rate; the operational target and an acceptable false-alarm rate remain to be defined and validated
User response time (median)	670 ms (successful only)	Hypothesis: end-to-end response time below the 670 ms unassisted median. The unassisted median already lies below 1000 ms, so a sub-second end-to-end target is not by itself an improvement and requires an explicit time origin
Response time variability	Very wide (max 23 613 ms)	Hypothesis: reduced dispersion of end-to-end response times; the dispersion metric and tolerance remain to be specified
System inference latency	N/A (no system)	Hypothesis: on-device inference with event-onset-to-alert latency small relative to the unassisted response benchmark; not yet measured
Sound classes supported	User perception only	Alarms, sirens, car horns, screaming and shouting, and dog barking
Notification channels	None	Screen, vibration, smartwatch, light signal

Note: KPI = key performance indicator; RT = reaction time.

Future validation should compare assisted and unassisted performance using predefined end-to-end latency, sensitivity, false alarm, missed detection, and usability outcomes. It should also use individually documented consent, participant-level audiological characterization, a fixed observation window, and a fully specified acoustic-calibration procedure.

CONCLUSIONS

This baseline study found a marked difference between the groups in both components of the behavioral response. The hearing-loss group had an overall no-response rate of 49.2% (52.3% within the standardized 5500 ms reference window), compared with 0% in the comparison group. Successful responses were also slower and more variable, with medians of 670 ms and 502 ms, respectively. These outcomes describe the recorded behavior in the experimental task. They should not be interpreted as direct evidence of whether a stimulus was consciously perceived or of the stage at which it was perceived.

The methodological contribution is an end-to-end evaluation framework that links a human behavioral baseline to acoustic-event detection and accessible notification. This framework extends conventional SED evaluation beyond classification metrics, but the present experiment did not test an AI model, notification interface, or assisted user response. Accordingly, the results support provisional design hypotheses rather than validated system requirements or claims of functional benefit. The time endpoints must remain distinct: event-onset-to-alert latency concerns the technical pipeline, alert-onset-to-user-response time concerns interaction with the notification, and end-to-end response time covers the complete interval from acoustic-event onset to user action.

The novelty of the proposed approach lies in integrating an empirical benchmark of hazardous sound responses for people with hearing loss, on-device detection, and a multi-channel notification layer. Its practical value will ultimately depend on demonstrated improvement in user safety and response behavior, not on internal classification performance alone. The current findings justify the next stage of development and evaluation, but do not establish the effectiveness of AI EAR SUPPORT itself.

ETHICS STATEMENT

The study was carried out as a non-medical, design-oriented measurement of behavioral response times, conducted to inform the development of an assistive system rather than to investigate health status, diagnose any condition, or evaluate any clinical intervention. No clinical, audiological, or demographic records were collected, processed, or stored beyond the group-classification variable required by the design, and no names or other direct identifiers were recorded at any point. The dataset consists of trial-level timing values and the corresponding sound-class labels, linked within a session by a randomly assigned study code so that observations belonging to the same participant could be grouped in the statistical analysis; no key permitting re-identification was created or retained. Participation was entirely voluntary, all participants were adults, and no procedure was invasive. The system was operated at a nominal presentation setting of approximately 75 dB SPL, which corresponds to an ordinary conversational-to-traffic listening level; because the presentation level was not verified with calibrated instrumentation, it is reported as a nominal setting rather than as a certified exposure value. Participants were free to discontinue at any time without consequence, and no participant reported discomfort or discontinued the session. Information about the purpose and course of the measurements was provided to participants in the hearing-loss group in Polish Sign Language by a licensed interpreter, and organizational arrangements were made jointly with the implementation partner, which is also the employer of the participating individuals. On this basis, and following consultation with the implementation partner, formal ethics committee approval was determined not to be required for this stage of the work. Consent to participate was documented at the group level rather than through individually signed forms. The authors recognize that individually documented informed consent and a formal institutional ethics determination represent better practice, and both will be obtained for the subsequent, system-validation studies in this research cycle, which will additionally involve participant-level audiological characterization.

Acknowledgements

The authors would like to thank the employees and management of IntroGraf (Lublin) for

their participation in the study and for the cooperative work environment that enabled the baseline measurements. The authors are particularly grateful to Magdalena Gach, licensed Polish Sign Language (PJM) interpreter, Vice-President of the "Migiem przez świat" association for deaf people, and appointed representative for deaf employees at IntroGraf – who coordinated the course of the experimental sessions and, without whose involvement, the measurements in the hearing-loss group would not have been possible. The doctoral research is conducted within the framework of the Doktorat Wdrożeniowy program of the Polish Ministry of Science and Higher Education, in collaboration with the implementation partner Freshware Sp. z o.o.

Publication costs for this article were covered under the Doktorat Wdrożeniowy agreement funded by the Polish Ministry of Science and Higher Education (agreement No. DWD/8/0285/2024).

REFERENCES

1. Jia W., Shi L. The Effect of Auditory Stimuli in Audio-Visual Two-Source Integration. 2019 11th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), Hangzhou, China, 2019, pp. 145–148. https://doi.org/10.1109/IHMSC.2019.00041
2. Kaufmann T.B., Foroogozar M., Liss J., Berisha V. Requirements For Mass Adoption Of Assistive Listening Technology By The General Public. 2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW), Rhodes Island, Greece, 2023, pp. 1–5. https://doi.org/10.1109/ICASSPW59220.2023.10193566
3. Williams J., Azim T., Piskopani A.-M., Chamberlain A., Zhang S. Socio-Technical Trust For Multi-Modal Hearing Assistive Technology. 2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW), Rhodes Island, Greece, 2023, pp. 1–5. https://doi.org/10.1109/ICASSPW59220.2023.10193586
4. Xiao Y., Yin H., Bai J., Das R.K. DG-SED: Domain Generalization for Sound Event Detection with Heterogeneous Training Data. 2025 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Singapore, Singapore, 2025, pp. 143–148. https://doi.org/10.1109/APSIPAASC65261.2025.11249270
5. Hu J. et al. PSELDNets: Pre-trained neural networks on a large-scale synthetic dataset for sound event localization and detection. IEEE Trans. Audio Speech Lang. Process. 2025; 33: 2845–2860. https://doi.org/10.1109/TASLPRO.2025.3587446
6. Li Y., Wang X., Liu H., Tao R., Yan L., Ouchi K. Semi-Supervised Sound Event Detection with Local and Global Consistency Regularization. ICASSP 2024 – 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Seoul, Korea, 2024, pp. 271–275. https://doi.org/10.1109/ICASSP48485.2024.10446386
7. Woods D.L., Wyma J.M., Yund E.W., Herron T.J., Reed B. Factors Influencing the Latency of Simple Reaction Time. Front. Hum. Neurosci. 2015; 9: 131. https://doi.org/10.3389/fnhum.2015.00131
8. Zhang B., de Winter J.C.F., Varotto S., Happee R., Martens M. Determinants of take-over time from automated driving: A meta-analysis of 129 studies. Transp. Res. Part F Traffic Psychol. Behav. 2019; 64: 285–307. https://doi.org/10.1016/j.trf.2019.04.020
9. Weaver B.W., DeLucia P.R. A Systematic review and meta-analysis of takeover performance during conditionally automated driving. Hum. Factors 2022; 64(7): 1227–1260. https://doi.org/10.1177/0018720820976476
10. Bottari D., Caclin A., Giard M.H., Pavani F. Changes in early cortical visual processing predict enhanced reactivity in deaf individuals. PLoS ONE 2011; 6(9): e25607. https://doi.org/10.1371/journal.pone.0025607
11. Gioiosa Maurno N., Phillips-Silver J., Daza González M.T. Research of visual attention networks in deaf individuals: A systematic review. Front. Psychol. 2024; 15: 1369941. https://doi.org/10.3389/fpsyg.2024.1369941
12. Retter T.L., Webster M.A., Jiang F. Directional visual motion is represented in the auditory and association cortices of early deaf individuals. J. Cogn. Neurosci. 2019; 31(8): 1126–1140. https://doi.org/10.1162/jocn_a_01378
13. Albash N.I.A. Looking ahead: A systematic review of the research on AI-based applications for the deaf and hard of hearing (2000–2024). Int. J. Innov. Res. Sci. Stud. 2025; 8(6): 2527–2537. https://doi.org/10.53894/ijirss.v8i6.10136
14. Cakır E., Parascandolo G., Heittola T., Huttunen H., Virtanen T. Convolutional recurrent neural networks for polyphonic sound event detection. IEEE/ACM Trans. Audio Speech Lang. Process. 2017; 25(6): 1291–1303. https://doi.org/10.1109/TASLP.2017.2690575
15. Gemmeke J.F., Ellis D.P.W., Freedman D., Jansen A., Lawrence W., Moore R.C., Plakal M., Ritter M. Audio Set: An Ontology and Human-Labeled Dataset for Audio Events. 2017 IEEE

International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 2017, pp. 776–780. https://doi.org/10.1109/ICASSP.2017.7952261

16. Salamon J., MacConnell D., Cartwright M., Li P., Bello J.P. Scaper: A Library for Soundscape Synthesis and Augmentation. 2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), New Paltz, NY, USA, 2017, pp. 344–348. https://doi.org/10.1109/WASPAA.2017.8170052
17. Yin H., Xiao Y., Bai J., Das R.K. Leveraging LLM and Text-Queried Separation for Noise-Robust Sound Event Detection. 2025 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW), Hyderabad, India, 2025, pp. 1–5. https://doi.org/10.1109/ICASSPW65056.2025.11011161
18. Khandelwal T., Das R.K., Chng E.S. Is Your Baby Fine at Home? Baby Cry Sound Detection in Domestic Environments. 2022 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Chiang Mai, Thailand, 2022, pp. 275–280. https://doi.org/10.23919/APSIPAASC55919.2022.9980350
19. Fu G.-S., Senechal T., Challenner A., Zhang T. Unified Speculation, Detection, and Verification Keyword Spotting. ICASSP 2022 – 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, Singapore, 2022, pp. 7557–7561. https://doi.org/10.1109/ICASSP43922.2022.9747243
20. Behnam P. et al. Hardware–software co-design for real-time latency–accuracy navigation in tiny machine learning applications. IEEE Micro 2023; 43(6): 93–101. https://doi.org/10.1109/MM.2023.3317243
21. Tonami N., Imoto K., Okamoto Y., Fukumori T., Yamashita Y. Sound Event Detection Based on Curriculum Learning Considering Learning Difficulty of Events. ICASSP 2021 – 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 2021, pp. 875–879. https://doi.org/10.1109/ICASSP39728.2021.9414184
22. Carmel D., Yeshurun A., Moshe Y. Detection of Alarm Sounds in Noisy Environments. 2017 25th European Signal Processing Conference (EUSIPCO), Kos, Greece, 2017, pp. 1839–1843. https://doi.org/10.23919/EUSIPCO.2017.8081527
23. Paus A., Große-Kampmann M., Bök P.-B. Performance Evaluation of AI-based Live Siren Detection on Smartphones in Vehicular Scenarios. 2026 International Wireless Communications and Mobile Computing (IWCMC), Wuzhou, China, 2026, pp. 1290–1296. https://doi.org/10.1109/IWCMC69287.2026.11580044
24. Guo T. et al. A Basic Study on Capacity and Reaction Time of Visual Working Memory for Elderly Memory Training. 2019 IEEE International Conference on Mechatronics and Automation (ICMA), Tianjin, China, 2019, pp. 600–604. https://doi.org/10.1109/ICMA.2019.8816525
25. Wu Y. et al. Exploring the Impact of Auditory Cues on Dynamic Balance Performance in Virtual Reality: A Comparative Study of Headphones and Loudspeakers. 2024 IEEE International Conference on Artificial Intelligence and eXtended and Virtual Reality (AIxVR), Los Angeles, CA, USA, 2024, pp. 384–391. https://doi.org/10.1109/AIxVR59861.2024.00065
26. Keerthika N., Sathish E., Kiruthika V., Santhakumar M. A Novel Design for Automatic Measurement of Reaction Time for Audiovisual and Muscular Stimulus. 2023 International Conference on Bio Signals, Images, and Instrumentation (ICBSII), Chennai, India, 2023, pp. 1–6. https://doi.org/10.1109/ICBSII58188.2023.10181095
27. Envato Elements. Sound Effects Library. Envato Pty Ltd, Melbourne, Australia. https://elements.envato.com/sound-effects (accessed 13 August 2026).
28. Madison G., Woodley of Menie M.A., Sänger J. Secular slowing of auditory simple reaction time in Sweden (1959–1985). Front. Hum. Neurosci. 2016; 10: 407. https://doi.org/10.3389/fnhum.2016.00407
29. Zhu K. et al. The Effects of Spatially-Informed Alarm Types and Presentation Rates on Pilot Performance. 2023 29th International Conference on Mechatronics and Machine Vision in Practice (M2VIP), Queenstown, New Zealand, 2023, pp. 1–5. https://doi.org/10.1109/M2VIP58386.2023.10413403
30. Hershey S., Chaudhuri S., Ellis D.P.W., Gemmeke J.F., Jansen A., Moore R.C., Plakal M., Platt D., Saurous R.A., Seybold B., Slaney M., Weiss R.J., Wilson K. CNN Architectures for Large-Scale Audio Classification. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 2017, pp. 131–135. https://doi.org/10.1109/ICASSP.2017.7952132
31. Kong Q., Cao Y., Iqbal T., Wang Y., Wang W., Plumbley M.D. PANNs: Large-Scale Pretrained Audio Neural Networks for Audio Pattern Recognition. IEEE/ACM Trans. Audio Speech Lang. Process. 2020; 28: 2880–2894. https://doi.org/10.1109/TASLP.2020.3030497
32. Choudhary S., Karthik C.R., Lakshmi P.S., Kumar S. LEAN: Light and Efficient Audio Classification Network. 2022 IEEE 19th India Council International Conference (INDICON), Kochi, India, 2022, pp. 1–6. https://doi.org/10.1109/INDICON56171.2022.10039921

33. Magno M., Mayer P., Benini L. A Self-Sustaining Micro-Watt Programmable Smart Audio Sensor for Always-On Sensing. 2018 Ninth International Green and Sustainable Computing Conference (IGSC), Pittsburgh, PA, USA, 2018, pp. 1–8. https://doi.org/10.1109/IGCC.2018.8752147
34. Wang Y., Habib A. Protect Data Confidentiality for on-Device Machine Learning Through Split Inference. 2025 10th International Conference on Fog and Mobile Edge Computing (FMEC), Tampa, FL, USA, 2025, pp. 290–297. https://doi.org/10.1109/FMEC65595.2025.11119383
35. Badiger S., Sandipkumar A., Shaikh A., Jaiswal A.B. Edge-AI Based Activity Recognition for Elderly Homes. 2026 International Conference on Signal, Systems, and Computing for Next-Gen Automation (ICSSCNA), Eluru, India, 2026, pp. 1210–1215. https://doi.org/10.1109/ICSSCNA68616.2026.11546712
36. Rurenko A., Anuragi D., Farooq A., Salmimaa M., Radivojevic Z., Kumpulainen S., Raisamo R. Sensing what you do not see: alerting of approaching objects with a haptic vest. Sensors 2025; 25(18): 5808. https://doi.org/10.3390/s25185808
37. Blum J.R. et al. Getting your hands dirty outside the lab: A practical primer for conducting wearable vibrotactile haptics research. IEEE Trans. Haptics 2019; 12(3): 232–246. https://doi.org/10.1109/TOH.2019.2930608