forum acusticum 2026
8-12 September · Graz, Austria

Sound Source Localization: a Study comparing Angular and Distance Perception Accuracy in a Vehicle Cabin and in a Reference Room

Alessandro Travaglini, Álvaro Jesús Nieto Illescas, Luis Martínez, Johannes Klein

Harman Becker Automotive Systems GmbH

Abstract

This paper compares sound source localization performance in a vehicle and in a multi-channel reference room (ITU-R BS.775-4 [1] standard), to assess whether the latter can be a suitable environment for preliminary tuning and verification of spatial localization behavior. The same listening tests (angular localization and distance-movement identification) were run with 16 listeners in both environments, using two stimulus classes: a female navigation-speech prompt and a Park Distance Control (PDC) warning chirp. Angular localization was of a comparable order in the two environments: about 77% of judgments in the room and 67% in the car fell within ±20 degrees of the true direction, with almost identical median errors; the superior performance in the room came mainly from fewer gross mislocalizations. The car additionally showed a front-left directional bias linked to a non-center listening position and to the different geometry of the speaker layout. The distance perception testing showed inconsistent performance in both room and car: near-field away movements were frequently confused with "steady", while far-field away movements were identified very accurately. For the away case, detection generally improved as the movement occurred farther from the listener.

Keywords: *sound source localization, spatial audio, vehicle cabin, reference listening room, distance perception*

1. Introduction

Accurate spatial audio rendering in vehicles is essential for applications such as advanced driver-assistance alerts and immersive audio experiences. Placing warning sounds or navigation cues at precise virtual locations can intuitively guide driver’s attention. However, achieving high localization accuracy in a car cabin is challenging due to the acoustic environment and speaker layout. Unlike an ideal audio room, where listeners sit equidistant from symmetrically placed full-range speakers in a controlled acoustic space, a vehicle interior is a small, asymmetric environment. Loudspeakers are positioned based on interior design constraints (e.g., in doors, dashboard, rear deck), causing irregular distances and angles relative to the listener. Moreover, car speakers often cover limited frequency bands (woofers, midranges, tweeters), meaning a single sound's spectral content is split across multiple locations.

The objective of this study is to quantitatively assess localization accuracy in a vehicle cabin, to compare it directly against the same test in a multi-channel reference listening room. We focus on two aspects of localization: directional (angular) accuracy and distance perception, and we examine how source angle, source content, and environment influence performance.

2. Methodology

2.1 Test Environments

The test was arranged as two independent sessions in the following environments.

2.1.1 In-Car (Vehicle Cabin)

A test vehicle located in HARMAN offices in Garching (Germany) was outfitted with an automotive audio system running a proprietary spatial rendering technology that reproduces intended source directions and distances over the vehicle's loudspeaker layout.

The car audio system comprises a 16-loudspeaker architecture distributed throughout the cabin. The front stage consists of two 25 mm tweeters, two 100 mm midrange drivers, and two 170 mm door woofers, complemented by a 90 mm center-channel mid-tweeter

Corresponding author*: Alessandro.traglini@harman.com*

speaker. The rear section includes two 90 mm midrange/tweeter units and two 170 mm woofers in the rear doors, together with two additional 90 mm surround loudspeakers located behind the rear seats.
The system, including the spatial rendering technology, was tuned in-situ for optimizing the vehicle's cabin acoustics and signed off by experts in audio perception. The driver's seat was set as the listening position (left-hand-drive vehicle), and cabin ambient noise was minimized for testing.

2.1.2 Reference Multichannel Listening Room

An acoustically treated audio laboratory, located in HARMAN offices in Queretaro (Mexico), meeting ITU-R BS.775-4 [1] requirements for 7.1 reproduction system loudspeaker arrangement, seven full-range loudspeakers arrayed in a circle around the listener at standard angles.
The room is acoustically treated to minimize reflections, providing a neutral environment for sound localization. The same test signals used in the car were reproduced through the corresponding loudspeakers.

2.2 Stimuli

The investigation was designed to examine practical use cases and consequently incorporated two distinct classes of stimuli aligned to -16.0 LUFS (per ITU-R BS.1770 [2])

Speech Prompt (Navigation Voice): a neutral female voice providing navigation instructions.

PDC Warning Chirp: a sequence of 5 short beeps (100 ms each, spaced 250 ms apart), fundamental at around 500 Hz with harmonics up to 5th order, including both even and odd components, monotonically decreasing modal gain.

2.3 Participants

The experiment involved 32 subjects across two locations, with 16 participants in each environment, both male and female, aged between 23 and 53, all employees of HARMAN International. All participants reported having no hearing impairment and were trained in the test procedure prior to the task.

2.4 Procedure and Tasks

Each listener completed the same two main tasks, either in the car or in the reference room.

Angle Localization Task: a single sound item (speech phrase or PDC chirp sequence) was played from a virtual direction. The listener was tasked to indicate the perceived direction by selecting the perceived position within two concentric rings on a GUI representation of the horizontal plane (Figure 1). The interface intentionally accepts any click radius in that ring and uses only the click direction to score angular perception. After selecting the perceived location, the subjects were prompted to confirm the selection is valid or repeat the stimulus reproduction.

Distance Movement Task: three consecutive sound items were played from the same angular direction, with the virtual distance of the source either decreasing, increasing, or unchanged over the sequence. The listener's task was to identify the change in distance ("moving closer", "moving away", or "no change"). Trials included different angles,

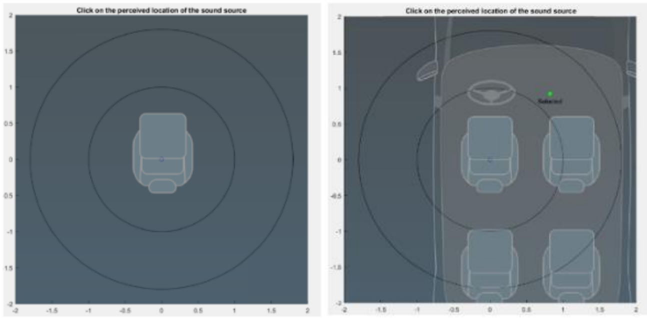

Figure 1. Response interface and environment layout: the reference room (left) and the vehicle cabin (right).

initial distances (Near vs. Far) and movement types, randomized across both content types. Each moving sequence used a fixed step size: the standard conditions shift the source by 20 percentual units of range per step (a 40-unit total span). One additional condition, Middle Away**, deliberately used a smaller 10-unit step (a 20-unit span) to probe sensitivity to a half-magnitude movement.
Source distance is expressed as a 0 to 100 units scale of the virtual distance range resulting from the rendering of the distance stimuli, where 0 is closest to the listener and 100 is farthest, so an increasing sequence is a source moving "away" and a decreasing sequence a source moving "closer". Table 1 lists the full set of conditions and their distance sequences; each was presented for both content types (PDC chirp and navigation speech) at five source directions.

Table 1. Distance test movement conditions. The seven primary conditions move the source in equal 20-unit steps (a 40-unit span); the Middle Away** condition moves in half-size 10-unit steps (a 20-unit span) and is analyzed separately

Condition	Distance sequence
Near Away	0 → 20 → 40
Middle Away	30 → 50 → 70
Far Away	60 → 80 → 100
Far Closer	100 → 80 → 60
Middle Closer	70 → 50 → 30
Near Closer	40 → 20 → 0
Middle Steady	50 → 50 → 50
Middle Away**	40 → 50 → 60

The diagram in Figure 2 shows the virtual locations of the test cases used in angle and distance listening tasks. The vehicle top-view is shown as reference. Purple markers indicate the expected angle-of-arrival positions for the angle-localization stimuli, all presented at the closest virtual distance. Colored three-point trajectories depict the distance-evolution stimuli used in the distance movement task, each trajectory reproduced during the task is drawn (-30, 150, 180, -90 and 90 degrees) and represents a condition sequence as shown in Table 1 (away, closer, or steady). Arrowheads indicate motion direction. The blue marker at the center indicates the listener position.

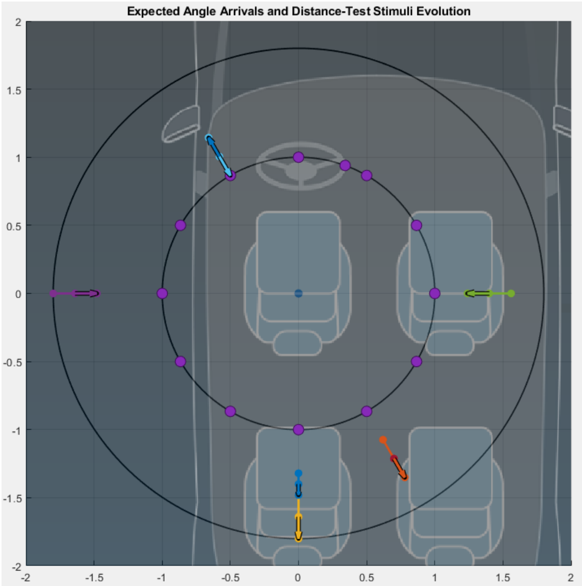

Figure 2. Visualization of all presented test cases. Purple dots: angle test, Arrows: Distance Movement

The data collected in the listening assessments included angular error per trial (actual vs. perceived angle) and accuracy of movement direction recognition. Before the execution of the formal test, a short training session based on the same UI, playing stimuli in different angles and distances was performed to help listeners adapting to the virtual sound movements and direction of arrival.

3. Angular Localization Results

3.1 Overall Accuracy

Angular accuracy was of a comparable order in the two environments. In the reference room, about 77% of listener judgments fell within ±20 degrees of the true source direction; in the car, about 67% did.

Median absolute error was almost identical in the two environments (about 11 degrees in the room and 12 degrees in the car), i.e. half of the responses fell within roughly a dozen degrees of the true direction in both cases. The difference between environments was therefore driven less by the typical response than specific cases: the car produced roughly twice as many gross mislocalizations, which raised its mean error (about 20 degrees, versus about 15 degrees in the room) while leaving the median close to the room's.

Figures 3 and 4 show the median regression results for the reference room and vehicle cabin, respectively. The green dots are PDC stimuli and red dots are Navigation speech.

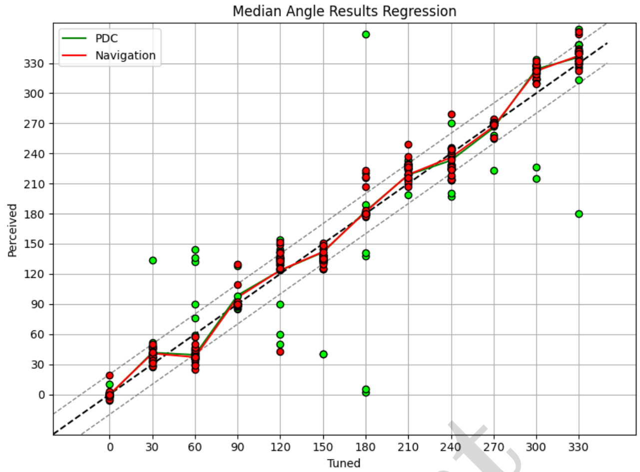

Figure 3. Angle detection: median regression / diagonal distribution, reference room

Broken down by content, speech was localized more accurately than the PDC chirp in both environments (room: 81% vs 73% within ±20 degrees; car: 72% vs 62%), possibly due to the broader spectral content of speech providing richer directional cues.

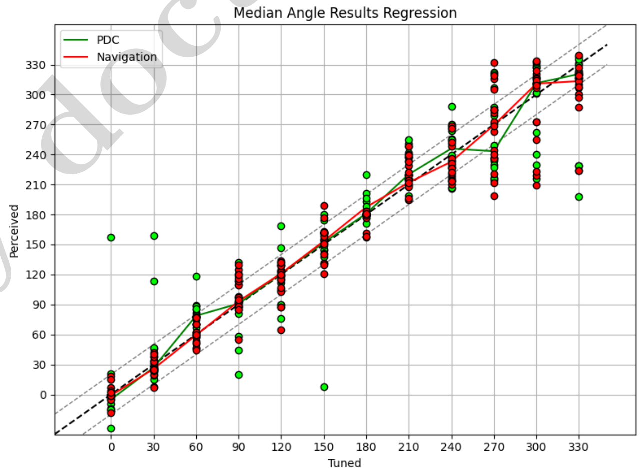

Figure 4. Angle detection: median regression / diagonal distribution, vehicle cabin

Beyond the average error, it is worth asking what drives the car's extra uncertainty: a single directional flaw shared by every listener, or larger deviation across individual listeners. The evidence points mainly to the latter. Listeners agreed with one another far less in the car: for a given target direction their answers were spread about 27 degrees apart, versus 19 in the room, and a standard agreement score (ICC, on a 0-to-1 scale where 1 is perfect agreement) fell from 0.24 in the room to just 0.07 in the car. In other words, the same virtual direction was heard quite differently from one person to the next in the car — most plausibly because individual head and ear geometry, exact seating relative to the asymmetric speakers, and listening skill matter much more there.

Table 2. Summary of angular test results accuracy

	Room	Vehicle
Angle within ±20 deg	77%	67%
Angle median |error|	11 deg	12 deg
Angle mean |error|	15 deg	20 deg
Angle within ±20 deg	81% / 73%	72% / 62%

3.2 Systematic Biases and Angle Dependency

The car showed a directional bias which was not present in the room. For sources in the front-left of the driver (around -30 to 0 degrees), perceived angles were pushed further to the left than intended, by 12 to 28 degrees on average (largest at about -30 degrees). This bias is attributed to the proximity of the front-left woofer, midrange and tweeter close to the driver, which can dominate the cues from other speakers. This bias was not observed in the reference room, whose symmetrical speaker placement produced small, non-directional errors across the front hemisphere.

However, it seems that the fact that subjects could see a physical loudspeaker in the room led to a consistent bias in the perception: particularly when the source was expected to arrive at direction of 60º (between two physical loudspeakers) most subjects reported it coming from directions closer to the front right loudspeaker placed at 30º.

Aside from this bias, angular performance was otherwise consistent across directions in both environments. Listeners clearly distinguished left from right and had little difficulty at front center (0 degrees) or far lateral positions (about ±90 degrees). Even when errors occurred, they tended to be moderate misjudgments rather than wrong-side responses; there was no indication of random guessing. Front/back discrimination was not explicitly tested; however, in general, listeners can experience front/back confusion in symmetric environments without visual cues, as can be observed in Figure 3 for the 180º tuned PDC content [3].

Absolute error did vary significantly test as One-way repeated-measures ANOVA results in $p < 0.001$ in all listening tests except for the PDC stimuli in the car where $p = 0.106$.

4. Distance Perception Results

4.1 Overall Accuracy

Distance-movement identification showed the lowest detection accuracy of the two tasks in both environments.

Scoring of perceived movement resulting from the distance testing over the 7 primary 20-unit conditions (half-magnitude test case Middle Away** probe excluded), is depicted in Figure 5, respectively for the reference room (top) and the vehicle cabin (bottom).

The most common error was generated when detecting a closer or away movement as "steady" (i.e. signaling “steady” a sound which was actually moving closer or away); false alarms (reporting a steady stimulus as moving) were also observed, although less frequently.

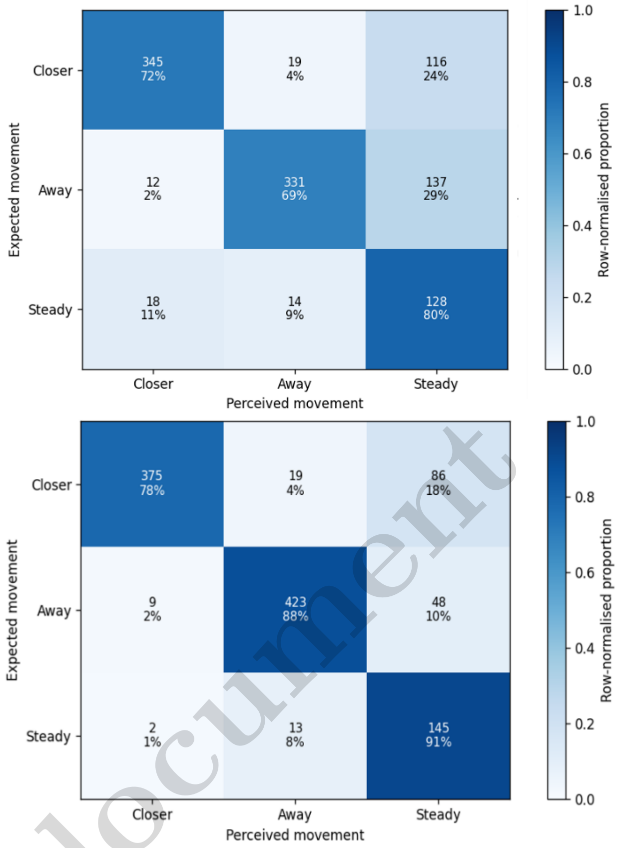

Figure 5. Distance-movement confusion matrix, reference room (top) vs vehicle cabin (bottom); rows = actual movement, columns = perceived movement; row-normalized. The dominant off-diagonal cell in both environments is away/closer heard as "steady".

4.2 Effects of Distance Range and Direction

The most consistent finding of the distance test was that the ability to perceive a change depended strongly on how far from the listener the movement occurred.

The clearest case is the "away" movement. In both environments, a source moving away (direction of movement away from the listening position) in the near field was often not reported as moving away at all: in the room it was judged "steady" more often than "away" (54% vs 42%), and in the car it was still confused with "steady" in about one trial out of five. As the same movement was placed farther from the listener (far field), away-detection scored about 97-99% in both environments. The trend was common to room and car, with the latter enabling a better distance perception.

For "closer" movements (direction towards the listening position) the field dependency was weaker and less consistent across environments (near/mid/far “closer”: room 68/62/86 %; car 92/78/65 %): the scoring trend even reverses, showing better scores in the far field in the room, and better scores in the near field in the car. This trend is most pronounced for the PDC chirp (near/mid/far “closer”: room 51 / 36 / 76, car 88 / 68 / 54), whereas navigation speech was detected more consistently across all fields in both environments (room 85 / 88 / 95, car 96 / 88 / 76).

The "steady" (no change in distance) case, tested at the midfield, was identified in 80% of trials in the room and 91% in the car.
Aggregated over the seven primary conditions (which all move the source in equal 20-unit steps) movement detection accuracy was 72% (room) and 84% (car).
Since the same rendered stimuli were presented in both environments, the fact that the near-field "away" ambiguity appears in the room as well as in the car identifies it as a perceptual effect rather than a vehicle-specific artefact.

Figure 6 lays out these responses in full: they show, for both reference room and car, a set of three heatmaps each one representing one of the three possible answers the listeners could give when presented with the distance stimulus: a) "Closer", b) "Away" and c) "Steady".
Within each heatmap the x-axis (columns) represents the eight field/movement conditions described in Table 1, in the order Near Away, Middle Away, Far Away, Far Closer, Middle Closer, Near Closer, Middle Steady and Middle Away**; the y-axis (rows) represents the ten stimuli of the two content types (PDC chirp and navigation speech) located at the five angles depicted in Figure 2 (-30, 150, 180, -90 and 90 degrees).
Each cell visualizes the percentage of detections of the test item indicated on the Y axis signaled as the corresponding movement class of the Y axis. The color-coding ranges from 0% (dark) to 100% (bright).

4.3 Stimulus Content Influence

The analysis of the listening test results shows that the content had a modest effect overall for angular localization in both room and car (speech vs PDC within ±20 degrees: room 81% vs 73%, car 72% vs 62%) and for distance-movement detection in the car (88% vs 81%).
The one aspect content dependency clearly stood out was distance detection in the reference room, where speech was tracked far more reliably than the PDC chirp (82% vs 62% in the room compared to 88% vs 81% in the car).

It appears that the broadband speech signal provides richer distance and spectral cues than the narrowband, short length, high-frequency chirp; additionally, larger variability was observed in the room. In contrast, the vehicle cabin appears to provide additional auditory cues that improve movement detection and reduce the performance gap between speech and chirp stimuli.
The practical implication is that distance perception of a narrowband alert such as the PDC chirp is harder in an acoustically treated reference room than in a vehicle cabin.
All seven primary movement conditions shown in Table 1 were identified well above the one-third chance level in both environments (one-sided binomial tests, all $p < 0.001$); the only marginal case being near-field "away" in the room (42.5%, $p = 0.01$).

5. Discussion

The comparison shows that a well-calibrated vehicle cabin and a reference room deliver angular localization of a similar order (median error around a dozen degrees in both), with the room showing a slightly higher score.

The results underline that the error is likely due to the impact of asymmetric listening position and speaker placement: the proximity of one speaker to the listener, together with the routing of a single source to multiple speakers with different frequency ranges, can weaken the localization accuracy of phantom sources in the vehicle.
Distance perception proved to be more difficult than angular localization in both environments; importantly, both environments showed the same failure pattern. The near-field "away" movement being heard as "steady" is present in the room and in the car, indicating a perceptual limit rather than an environment-specific defect.
For applications this means that distance-based cues — and especially with narrowband alerts — should not be relied upon alone for critical information and may need to be exaggerated or complemented by visual or haptic feedback.

The test also highlighted that a reference room may not reproduce the vehicle cabin's absolute behavior (it is slightly better for angle detection and slightly worse for movement detection). However, it does reproduce some trends such as the content dependency and the near-to-far movement detection.
A reference room is therefore a reasonable environment for preliminary localization tuning and verification, provided its figures are treated as indicative and the vehicle remains the target acoustic environment for validating automotive audio technologies, particularly for evaluating placement-dependent biases.
Halving the movement step-size (stimuli depicted in row 8 of Table 1) roughly halved detection accuracy in both environments ("Middle Away**" 38% room, 45% car, versus the 20 unit step "Middle Away" 68% and 87%), where the errors were due to movements being heard almost entirely as "steady" (61% room, 53% car, versus 29% and 11% for the full-size movement).
Since the smaller movement was often reported as "no change" in both environments, it suggests a magnitude factor rather than a room-vs-car difference.

These observations are consistent with the spatial-audio literature. Blauert (1997) [4] emphasizes the perceptual strength of azimuth cues over distance cues, matching the angle-vs-distance gap reported in this paper. Begault (1994) [5] documents the difficulty of reproducing accurate distance cues in virtual settings, similarly to the field-dependent distance results we collected.

6. Conclusion

Running identical angular and distance-movement localization tests in a vehicle cabin and in an ITU-R BS.775-4 [1] reference room shows that the two environments agree in trend more than in absolute behavior.
Angular localization was comparable (about 77% within ±20 degrees in the room and 67% in the car, median error 11 vs 12 degrees), with the room's deviation coming from fewer gross errors. The car showed an additional front-left bias from asymmetric listening position.
Distance-movement identification was fairly similar in both environments (72% room, 84% car): near-field away movements were frequently confused with "steady", while

far-field away movements were identified almost perfectly. As the pattern appeared in both the room and vehicle cabin, it is likely to reflect a general characteristic of auditory distance perception rather than a vehicle, content, or virtual technology specific effect.

In conclusion, in-car spatial audio can convey direction reliably enough to be useful for driver alerts and augmented reality applications; on the contrary, conveying of distance information via sound cues shows to be unreliable and content dependent; moreover, the results suggest that a reference room can provide useful preliminary indications of localization performance, especially for navigation speech or broadband complex content rather than warning chirps, although final validation remains necessary in the target vehicle environment for evaluating absolute performance and placement-specific biases.

References

[1] ITU-R BS.775-4. Multichannel stereophonic sound system with and without accompanying picture.

[2] ITU-R BS.1770. Algorithms to measure audio programme loudness and true-peak audio level..

[3] Moore, Brian C.J. (2004). An Introduction to the Psychology of Hearing. Elsevier.

[4] Blauert, J. (1997). Spatial Hearing: The Psychophysics of Human Sound Localization. MIT Press.

[5] Begault, D. R. (1994). 3-D Sound for Virtual Reality and Multimedia. Academic Press

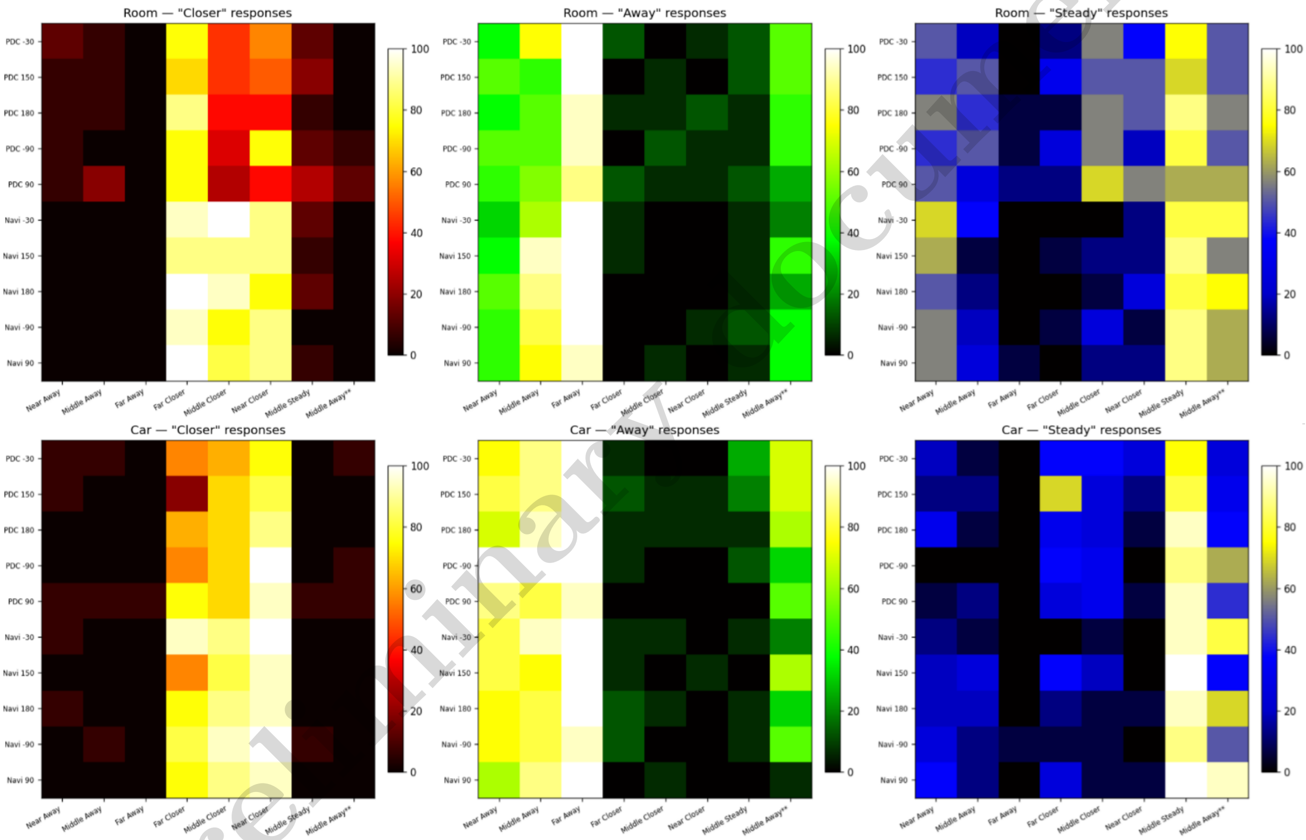

Figure 6. Distance-response heatmaps for reference room (top) and vehicle cabin (bottom), percentage of:
a) "Closer" / b) "Away" / c) "Steady" responses for each field/movement condition (columns), broken down by content and source angle.