

Article

Not peer-reviewed version

Effect of Edge-Based Acoustic Modifications on Speech Intelligibility and SNR Thresholds in Classrooms: A Field Study

[Sebastian Kümritz](#)*

Posted Date: 31 March 2026

doi: 10.20944/preprints202603.2422.v1

Keywords: classroom acoustics; sound field diffuseness; speech intelligibility; edge-based acoustic treatments



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Effect of Edge-Based Acoustic Modifications on Speech Intelligibility and SNR Thresholds in Classrooms: A Field Study

Sebastian Kümritz

H2 Think gGmbH, Germany; kuenmritz@h2think.org

Abstract

Classroom acoustic design is typically guided by reverberation-time-based criteria, assuming that global decay adequately predicts speech intelligibility. This field study examines whether edge-based acoustic treatments influence speech intelligibility thresholds under steady background noise beyond conventional decay metrics. Four geometrically comparable primary-school classrooms were investigated: one acoustically optimized ceiling reference, one untreated control, and two rooms equipped with edge-based systems (primarily absorptive vs. primarily reflective/scattering). Room impulse responses were measured at multiple source–receiver configurations, and standard parameters (STI, C_{50} , D_{50} , T_{30}) were derived. Psychoacoustic sentence-repetition tests with children (7–9 years) under calibrated noise conditions were used to determine speech-level and SNR thresholds at a predefined sentence-repetition intelligibility criterion. Both edge-based treatments improved STI, C_{50} , and D_{50} relative to the untreated room, while affecting T_{30} differently. In the psychoacoustic evaluation, untreated conditions required the highest SNR thresholds. Both treated rooms showed reduced SNR requirements compared to the untreated state. Although not all SNR comparisons reached statistical significance after correction for multiple testing, effect sizes indicated consistent directional improvements. This study was designed as an exploratory pilot investigation to assess the plausibility and directionality of potential effects rather than to provide definitive statistical confirmation. The findings suggest that intelligibility under noise is not fully explained by reverberation time alone and motivate further investigation of structural room-acoustic characteristics in small classrooms.

Keywords: classroom acoustics; sound field diffuseness; speech intelligibility; edge-based acoustic treatments

1. Introduction

Classroom acoustic quality has a direct and measurable impact on speech intelligibility and learning performance. Numerous experimental and review studies demonstrate that excessive reverberation and background noise impair speech perception, particularly in children, who exhibit reduced tolerance to degraded acoustic conditions compared to adults. Early work by Bradley et al [1] identified optimal reverberation times in classrooms of approximately 0.4–0.5 s for effective speech communication. Subsequent investigations by Dockrell and Shield [2] and Klatt et al [3] demonstrated that excessive reverberation and background noise impair attention, memory performance, and reading development in children.

More recently, systematic reviews have consolidated this evidence. Murgia et al [4] analyzed 23 classroom studies and reported consistent correlations between increased reverberation time and reduced speech intelligibility. Mealings and Buchholz [5] further emphasize that poor classroom acoustics disproportionately affect younger pupils and children with additional learning needs. These findings support the consensus that acoustic optimization is not merely a comfort measure but an educational necessity.

International standards reflect this evidence base. ISO 3382-2 [6] defines measurement procedures for reverberation time in ordinary rooms and is widely referenced for classroom assessment. National and international design guidelines typically recommend reverberation times between 0.4 s and 0.6 s for typical classroom volumes. DIN 18041 [7] specifies target values depending on room size and use in Germany, while ANSI/ASA S12.60 [8] defines maximum background noise levels and reverberation limits for U.S. classrooms. These standards are predominantly decay-time oriented and implicitly prioritize reverberation control as the primary predictor of speech clarity.

However, reverberation time alone does not fully describe speech-relevant room acoustics. Parameters such as C_{50} , D_{50} , and STI quantify the temporal distribution of sound energy and better reflect clarity-related properties. Vasconcelos Rabelo et al [9] demonstrated in a field study of 18 occupied classrooms that higher STI and shorter T_{30} were associated with improved student speech intelligibility. Arvidsson et al [10] showed experimentally that combinations of absorbers and diffusers allow targeted modification of C_{50} and sound strength, indicating that redistribution of early reflections can influence clarity independently of total energy reduction.

In parallel, research has increasingly addressed the role of room edges and boundary transitions. Kraxberger and Vorländer [11] demonstrated with a validated finite-element model that porous edge absorbers significantly influence modal decay and spatial energy distribution, while Kurz et al. [12] showed that strategically placed corner absorbers act as effective modal dampers. Diffusive treatments have likewise gained attention: Yoshida et al. [13] introduced crossed-rib diffusers enabling broadband scattering, Eldien and Hammad [14] reported STI improvements in classrooms using cymatic geometries, and Khrystoslavenko and Grubliauskas [15] experimentally quantified diffuser scattering coefficients, emphasizing the importance of controlled scattering for sound field uniformity.

Despite these findings, classroom design practice remains predominantly reverberation-time-centered. The spatial-temporal organization of early reflections has not been systematically operationalized as an independent performance parameter in classroom design practice. ISO 17497-1 [16] provides a framework for measuring scattering properties, yet no widely adopted descriptor translates diffuseness into classroom-specific design criteria.

This gap raises a fundamental question: can targeted edge-based acoustic interventions modify sound field structure in ways that measurably improve speech intelligibility under realistic noise conditions, beyond what is predicted by T_{30} alone? Put more concisely: what role does sound field structure play in speech intelligibility?

The present field study addresses this question in geometrically comparable primary-school classrooms. By combining spatially distributed room impulse response measurements, standardized acoustic parameter extraction (ISO 3382-1; IEC 60268-16 - [6,17]), and psychoacoustic intelligibility threshold testing under controlled noise, the study investigates whether edge-based absorptive and reflective/scattering systems influence speech perception beyond conventional decay metrics.

The findings aim to contribute to a broader refinement of current design practice: while reverberation time remains the primary control parameter for speech intelligibility, the results suggest that structural aspects of sound field organization—such as early reflection distribution and diffuseness—deserve consideration as complementary acoustic performance criteria.

The present study was conceived as an exploratory investigation to assess whether systematic, directionally consistent effects can be observed under realistic field conditions. The goal was to evaluate whether the observed trends indicate a meaningful need for further, more comprehensive research.

2. Materials and Methods

2.1. Rooms and Acoustic Treatments

The study was conducted in four geometrically comparable primary-school classrooms (R114–R117) of the "Storchengrundschule" in Schöneiche (b. Berlin), Germany. All rooms share a similar rectangular floor plan, ceiling height, and construction type. One exception is Room R117, which includes an

adjacent winter-garden structure with a larger glazed area. Despite this architectural difference, room volume and primary geometry were comparable to the other classrooms. A schematic floor plan of the rooms is shown in Figure 1.

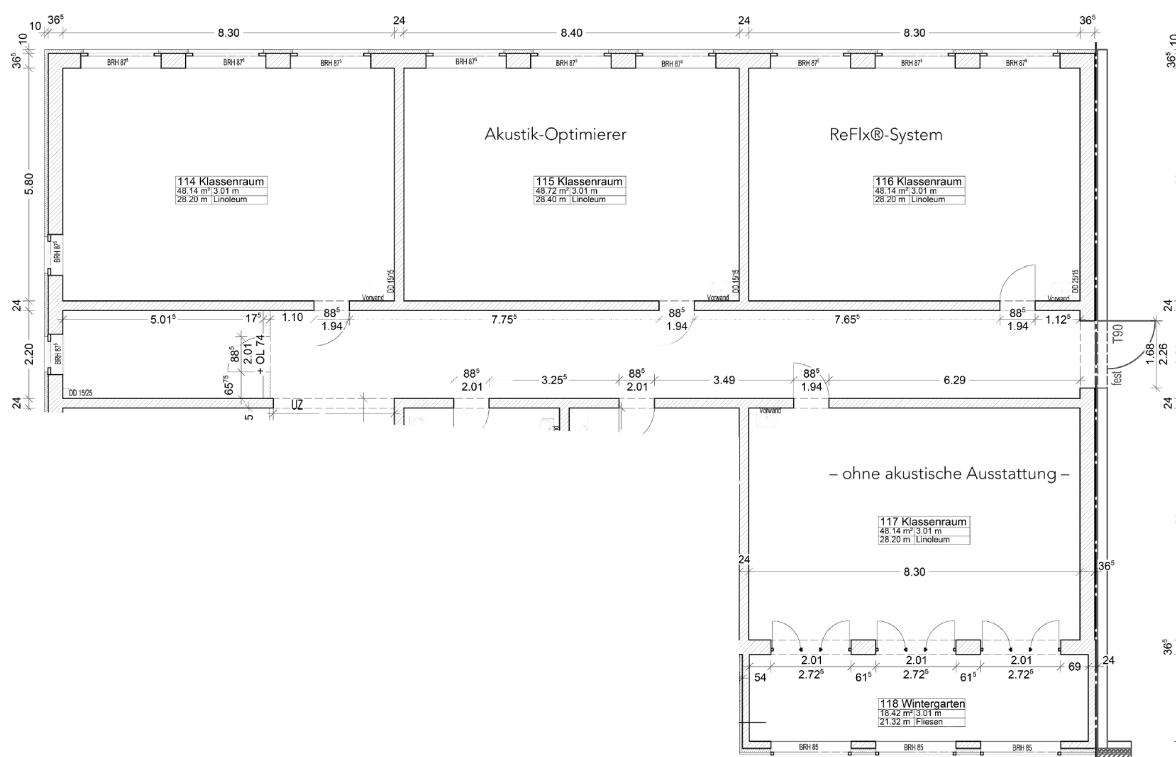


Figure 1. Schematic floor plan of the investigated classrooms (R114–R117).

Room R114 had previously been acoustically optimized using full ceiling absorption and served as a ceiling-absorption reference configuration. Rooms R115, R116, and R117 were initially untreated. During the study, R115 was equipped with an absorptive edge-based system (System 1), while R116 received a primarily reflective and scattering edge-based treatment (System 2). Room R117 remained untreated and served as a control.

System 1 (see Figure 2 left) consists of cylindrical acoustic elements positioned along room edges. The system primarily provides broadband sound absorption while introducing limited geometric scattering. The elements were manufactured by *EinrichtWerk GmbH*.

System 2 (see Figure 2 right) consists of inclined wooden panels installed at room edges, designed to redirect and scatter incident sound without introducing substantial additional absorption. The system (ReFlx[®]) was provided by *Raumakustik Premium e. K.*



Figure 2. Edge-based acoustic treatments investigated in the study. Left: System 1 (absorptive edge-based treatment). Right: System 2 (reflective/scattering edge-based treatment).

Both systems were installed at geometrically comparable edge positions and remained unchanged throughout the measurement period. The edge-based interventions investigated here are expected, based on their physical design, to modify the spatial redistribution of reflected sound energy. System 2 operates primarily through geometrically structured redirection and scattering at room edges, which can increase angular dispersion of reflected sound. System 1, although absorptive, selectively attenuates edge-related modal and grazing-incidence components while preserving ceiling-based reflection paths. Both mechanisms may influence the balance of reflected energy components within the early time window. The present investigation examines whether these edge-based modifications are reflected in conventional acoustic parameters (T_{30} , C_{50} , D_{50} , STI) and in psychoacoustic intelligibility thresholds.

2.2. Room Impulse Response Measurements

Room impulse responses (RIRs) were measured in all four classrooms using the exponential sine sweep method. As excitation signal, a logarithmic sine sweep covering the frequency range from 50 Hz to 20 kHz with a duration of 30 s was employed. The sweep was reproduced via a dodecahedral loudspeaker (Lookline SL 103AC, diameter 26 cm), positioned either at the teacher's desk (height $\sim 1,2$ m) or centrally in front of the blackboard (height $\sim 1,6$ m) to represent typical speech source locations.

The acoustic response was recorded using two measurement microphones (Behringer ECM 8000) connected to a two-channel USB audio interface (M-Audio Fast Track Pro). Recordings were performed with a sampling rate of 44.1 kHz. For each room, six receiver positions were defined, representing typical pupil seating areas and intermediate positions within the room. Measurements were conducted in pairs of spatially separated receiver positions (front, middle, or rear row), with two microphones recorded simultaneously in each configuration. Each source–receiver configuration was measured three times to reduce the influence of random disturbances, resulting in 36 measurements per room prior to data screening.

RIRs were calculated offline using a custom Python-based evaluation routine. The recorded sweep signals were temporally aligned and deconvolved in the frequency domain using the inverse filter derived from the time-reversed excitation signal. The resulting impulse responses were normalized and shifted such that the direct sound peak was located near the temporal origin. A time window of 4 s was retained for further analysis.

Following automated parameter extraction, all RIRs were visually inspected. Measurements showing ambiguous direct sound identification, elevated background noise, or clipping artefacts were excluded from further analysis. Depending on the room, between 0 and 6 measurements were removed.

Figure 3 illustrates representative impulse responses from Room 114. The right panel shows all measured RIRs of this room (grey), with one exemplary impulse response highlighted in blue. The displayed time range of 0–0.6 s visualizes the overall decay behaviour and reverberation characteristics. The left panel depicts the first 25 ms of the highlighted impulse response, allowing detailed inspection of the direct sound and prominent early reflections occurring at approximately 3.6 ms, 7.0 ms, and 9.5 ms.

For visualization, energy time curves (ETC) shown in Figure 3 were computed from the squared impulse response and smoothed (1 ms moving average) prior to logarithmic scaling. All acoustic parameters (T_{30} , C_{50} , D_{50} , STI) were calculated from the unsmoothed impulse responses.

A minimal reproducible example of the RIR processing workflow, including representative measurements from Room 114, is provided in [18].

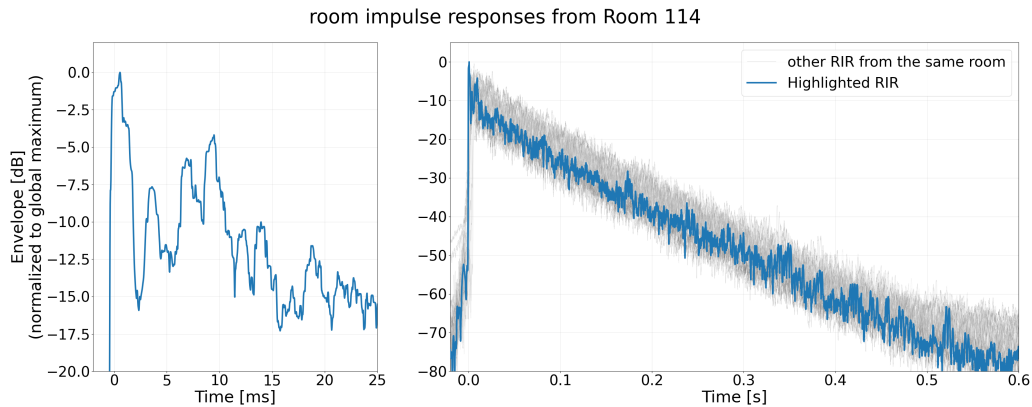


Figure 3. Room impulse responses measured in Room 114. Right: All measured RIRs (grey) with one representative response highlighted (blue), shown over a time range of 0–0.6 s. Left: Zoom into the first 25 ms of the highlighted RIR, illustrating the direct sound and early reflections at approximately 3.6 ms, 7.0 ms, and 9.5 ms.

2.3. Acoustic Parameter Calculation

All room acoustic parameters were derived from measured impulse responses in accordance with ISO 3382-1 [6]. The Speech Transmission Index (STI) was calculated according to IEC 60268-16 [17].

Reverberation Time T_{30}

Reverberation time T_{30} was derived from the backward-integrated energy decay curve (Schroeder method) by linear regression between -5 dB and -35 dB and extrapolated to a 60 dB decay:

$$T_{30} = 2 \cdot \Delta t \quad (1)$$

where Δt denotes the time interval between the decay curve crossing -5 dB and -35 dB of the decay curve.

Definition D_{50}

The definition describes the ratio of early sound energy within the first 50 ms to the total impulse response energy:

$$D_{50} = \frac{\int_0^{50 \text{ ms}} h^2(t) dt}{\int_0^{\infty} h^2(t) dt} \quad (2)$$

where $h(t)$ denotes the impulse response. Higher values of D_{50} are associated with improved speech intelligibility.

Clarity C_{50}

The clarity index expresses the ratio of early to late sound energy in logarithmic form:

$$C_{50} = 10 \log_{10} \left(\frac{\int_0^{50 \text{ ms}} h^2(t) dt}{\int_{50 \text{ ms}}^{\infty} h^2(t) dt} \right) \quad (3)$$

Positive values indicate a dominance of early reflections.

Speech Transmission Index STI

The Speech Transmission Index is based on the transmission of speech amplitude modulations through the acoustic channel. For each octave band k and modulation frequency m , a Modulation Transfer Index (MTI) is calculated:

$$\text{MTI}_{k,m} = \frac{m_{k,m,\text{out}}}{m_{k,m,\text{in}}} \quad (4)$$

The overall STI is obtained by weighted averaging across frequency bands and modulation frequencies. Values close to 1 indicate excellent speech intelligibility, whereas values below 0.3 correspond to poor conditions.

2.4. Spatial Coherence as a Proxy for Sound Field Diffuseness

To obtain an estimate of the spatial diffuseness of the sound field, a coherence-based proxy measure was evaluated from pairs of room impulse responses (RIRs) recorded at two spatially separated receiver positions.

The analysis was restricted to the late part of the impulse response in order to suppress the influence of direct sound and early reflections and to emphasize the reverberant sound field. Specifically, the coherence analysis was performed on time-windowed impulse responses, where only the late reverberant tail was considered. For each RIR $h(t)$, a time window $h(t_1 \leq t \leq t_2)$ was extracted, with $t_1 \approx 50$ ms (end of the early reflection regime) and $t_2 < T_{30}$, depending on the decay characteristics of the respective room. For each RIR pair $h_1(t)$ and $h_2(t)$, the magnitude-squared coherence function was computed in the frequency domain:

$$\gamma_{12}^2(f) = \frac{|S_{12}(f)|^2}{S_{11}(f) S_{22}(f)} \quad (5)$$

where $S_{12}(f)$ denotes the cross-power spectral density between the two signals and $S_{11}(f)$, $S_{22}(f)$ their respective auto-power spectral densities.

The coherence values were averaged over the frequency range from 250 Hz to 4000 Hz to obtain a single broadband descriptor for each measurement configuration. The calculations can be found in [18].

In a diffuse sound field, sound arrives from multiple directions with largely uncorrelated phase relationships, resulting in low spatial coherence between separated receiver positions. Conversely, higher coherence values indicate a more structured or directional sound field.

The resulting coherence values were therefore interpreted as a proxy for sound field diffuseness, where lower values correspond to a more diffuse field and higher values indicate reduced diffuseness. It should be noted that this measure does not represent a standardized diffuseness metric but serves as a comparative indicator within the present experimental framework.

2.5. Psychoacoustic Experimental Design and Procedure

The psychoacoustic experiment followed a within-subject design in which each participant (P_n) was tested in all four acoustic conditions. Measurements were conducted in two sequential blocks.

In block 1 (P_1 – P_4), all participants were tested consecutively within each room before the setup was relocated to the next room. The room order was 117→116→115→114. After a break, block 2 (P_5 – P_7) followed the reversed order 114→115→116→117. Within each block, participant order remained constant across rooms to ensure procedural consistency.

The study was conceived as an exploratory investigation, and the sample size was determined by practical feasibility within the school setting. It was considered sufficient to identify large directional effects, but not for confirmatory statistical inference.

Participants

A total of $N = 7$ children (3 female, 4 male; mean age 7.8 ± 0.8 years) participated in the study. All participants were primary-school pupils and native German speakers. No known hearing impairments were reported. Measurements were conducted during regular school supervision hours in coordination with educational staff. Participation was voluntary, and all recordings were anonymized prior to analysis.

Stimuli and Noise

Speech stimuli consisted of 41 semantically anomalous German sentences (e.g., “Der Teppich läuft nach Norden.” lit. “The carpet runs northwards”), designed to minimize semantic prediction and emphasize acoustically driven intelligibility. The stimuli were generated using a neural text-to-speech system with five virtual speakers (three female, two male voices).

Background noise was generated using the internal white-noise generator of the omnidirectional source (switch position “LINEAR”). The output level was set to -30 dB using the manufacturer-calibrated remote control. According to manufacturer specifications, these levels are internally referenced to the maximum output. Absolute sound pressure levels at the listening position were not measured; however, all playback settings and the electroacoustic chain remained identical across conditions, enabling consistent relative comparisons.

Procedure

Speech intelligibility was assessed using a sentence repetition task under continuous background noise. Participants were seated at a fixed listening position in the rear part of each classroom, while speech was presented from the front (teacher position) using a Bluetooth loudspeaker.

For each room condition, sentences were presented sequentially. The first sentence was played at maximum level; across subsequent trials, the speech level was reduced stepwise. If a sentence was not correctly reproduced, the level was maintained for the next trial. A run was terminated after repeated failures. Participant responses were given verbally immediately after each stimulus.

Two microphones were used for synchronous recording: a reference microphone placed near the loudspeaker to capture the presented stimulus, and a microphone at the listening position to record both the received speech signal and the participant’s verbal response.

The speech level $L_{S,room}$ represents a relative signal level (dBFS) derived from the recorded waveform and does not correspond to an absolute SPL. Consequently, reported values reflect relative signal levels rather than calibrated physical sound pressure levels, but allow consistent comparison across acoustic conditions.

2.6. Psychoacoustic Data Processing and Scoring

Recordings were processed offline. For each trial, the stereo recording was interpreted as reference channel (near the speech loudspeaker) and room channel (at the listener position). To focus on speech-relevant frequency content, signals were band-pass filtered (4th-order Butterworth, 300 Hz–3.4 kHz) using zero-phase forward–backward filtering. The effective stimulus window was determined automatically from the filtered clean stimulus using an envelope-based threshold criterion; this window was applied to the room-channel recording to estimate the speech signal level.

Background noise level at the listening position was estimated from two noise-only windows at the beginning and end of each recording (each approximately 0.45 s). The speech level $L_{S,room}$ was computed as the RMS level of the room-channel signal within the stimulus window (in dBFS). The noise level $L_{N,room}$ was computed analogously from the noise windows. The trial-specific SNR at the listening position was estimated by subtracting the noise power from the total signal power within the stimulus window and converting the resulting ratio to dB; negative or non-physical values were limited to -20 dB.

Speech intelligibility was scored from the participants’ verbal repetitions on a four-level ordinal scale: fully correct or clearly correct in meaning (1.0), minor errors but unambiguous sentence identification (0.7), fragments/isolated words only (0.3), and no intelligible reproduction (0.0). For each participant and room, an intelligibility threshold was derived for both $L_{S,room}$ and SNR_{room} by selecting trials with scores ≥ 0.7 (criterion: largely correct sentence reproduction) and calculating the 20th percentile of the corresponding level/SNR distribution.

2.7. ASR-Based Evaluation

In addition to human scoring, room-channel recordings were transcribed using the faster-whisper implementation of Whisper (large-v3 model). Transcription was performed with fixed language setting (German), beam size 5, and otherwise default inference parameters. Audio signals were resampled to 16 kHz prior to model inference.

The transcription and preprocessing workflow was implemented in Python and is publicly available in a dedicated repository [19].

For each trial, word-level agreement was calculated as the proportion of correctly transcribed words relative to the target sentence after removal of punctuation and capitalization differences. ASR-derived performance values were mapped onto the same ≥ 0.7 intelligibility criterion used in the human evaluation, and thresholds were derived using the identical 20th-percentile procedure.

2.8. Statistical Analysis

Statistical analyses were conducted separately for room-acoustic parameters derived from RIR measurements and psychoacoustic intelligibility thresholds.

Room-Acoustic Measurements

For each room and condition, 36 RIRs (12 source–receiver positions \times 3 repetitions) were evaluated. Acoustic parameters (T_{30} , C_{50} , D_{50} , STI) were derived from each RIR. As these measurements represent spatial samples of a single physical system rather than independent statistical observations, parameters are reported descriptively as medians with interquartile ranges (IQR).

Psychoacoustic Experiments

Seven participants were tested. Given the small sample size and potential deviations from normality, non-parametric methods were applied. Central tendencies are reported as medians with bootstrap-estimated 95 % confidence intervals (10,000 iterations)

Pairwise comparisons between acoustic conditions were conducted using the Wilcoxon signed-rank test for paired samples (two-sided). Effect sizes were calculated as

$$r = \frac{Z}{\sqrt{N}}, \quad (6)$$

where Z denotes the standardized test statistic and N the number of paired observations.

Overall agreement across multiple room conditions was evaluated using Kendall's coefficient of concordance (W).

Given the exploratory nature of the study and the limited sample size, statistical inference is interpreted cautiously, with emphasis placed on effect sizes and consistency of directional trends rather than solely on adjusted p-values.

3. Results

3.1. Room Acoustic Parameters

Table 1 summarizes STI, C_{50} , D_{50} , and T_{30} across rooms and treatment conditions. Parameters are reported as medians with interquartile ranges. The row Treat refers to the treatment (ceil = ceiling absorption; No = no treatment; S1 = System 1; S2 = System 2). The following comparisons refer primarily to median values.

The ceiling-absorption reference (R114) exhibited the most favorable room-acoustic conditions, with the highest median STI (0.51), C_{50} (+6.3 dB), and D_{50} (80.9 %), as well as the shortest reverberation time ($T_{30} = 0.48$ s). In contrast, the untreated rooms (R115, R116, and R117) showed consistently poorer values, with STI around 0.25, C_{50} near 0 dB or slightly below, D_{50} around 47–48 %, and T_{30} between 1.22 s and 1.32 s.

Table 1. Room-acoustic parameters across rooms and treatment conditions

Room	Treat	STI	C_{50} [dB]	D_{50} [%]	T_{30} [s]
114	Ceil	0.51 (0.49–0.54)	6.3 (5.9–7.3)	80.9 (79.7–84.3)	0.48 (0.47–0.48)
115	No	0.25 (0.24–0.26)	-0.6 (-0.8– -0.1)	46.6 (45.5–49.5)	1.22 (1.21–1.24)
115	S1	0.35 (0.34–0.38)	+2.6 (2.2–3.1)	64.3 (62.5–67.0)	0.75 (0.74–0.75)
116	No	0.25 (0.23–0.26)	+0.3 (-0.9–0.1)	48.3 (44.8–50.3)	1.32 (1.31–1.33)
116	S2	0.29 (0.29–0.32)	+0.9 (0.5–1.2)	55.0 (52.6–56.9)	1.02 (1.02–1.03)
117	No	0.25 (0.24–0.26)	-0.6 (-0.8– -0.1)	46.6 (45.5–49.5)	1.22 (1.21–1.24)

After installation of the edge-based treatments, all classical room-acoustic parameters improved. In R115, System 1 increased STI from 0.25 to 0.35, C_{50} from -0.6 dB to +2.6 dB, and D_{50} from 46.6 % to 64.3 %, while reducing T_{30} from 1.22 s to 0.75 s. In R116, System 2 also improved all parameters, though less strongly: STI increased from 0.25 to 0.29, C_{50} from +0.3 dB to +0.9 dB, and D_{50} from 48.3 % to 55.0 %, while T_{30} decreased from 1.32 s to 1.02 s.

Overall, the classical room-acoustic parameters followed a consistent pattern: the ceiling treatment produced the strongest improvement, System 1 yielded a substantial intermediate improvement, and System 2 a smaller but still noticeable one. Interquartile ranges were comparatively narrow across conditions, indicating only moderate spatial dispersion within rooms.

The spatial coherence analysis revealed clear differences between the investigated room conditions, as can be seen in Table 2. The ceiling-treated reference room (R114) exhibited substantially higher coherence values ($C_r = 0.31$) compared to all other conditions, indicating a more structured and less diffuse sound field.

Table 2. Spatial coherence C_r (250–4000 Hz) across rooms and treatment conditions. Values are reported as medians with interquartile ranges (IQR).

room	treatment	C_r	IQR
R114	ceiling	0.31	(0.29–0.32)
R115	no	0.12	(0.11–0.12)
R115	system1	0.19	(0.17–0.20)
R116	no	0.12	(0.12–0.12)
R116	system2	0.13	(0.13–0.14)
R117	no	0.12	(0.11–0.12)

Untreated rooms (R115, R116, R117) showed consistently low coherence values ($C_r \approx 0.11$ –0.12), suggesting a highly diffuse late sound field.

The edge-based treatments affected coherence differently. In R115, System 1 increased spatial coherence to $C_r = 0.19$, indicating a moderate reduction in diffuseness. In contrast, System 2 in R116 resulted in only a minor change ($C_r = 0.13$), remaining close to the untreated condition.

Interquartile ranges were small across all conditions, indicating low variability between measurement configurations and a high degree of spatial consistency within each room.

3.2. Psychoacoustic Intelligibility Thresholds

Figure 4 (left) shows the speech-level thresholds $L_{S,room}$ at which the predefined intelligibility criterion was reached. Values are reported as medians with 95 % bootstrap confidence intervals; individual data points are displayed.

Across conditions, the untreated room (R117) required the highest speech levels. All other rooms exhibited lower median thresholds. Pairwise Wilcoxon signed-rank tests (Holm-corrected) revealed statistically significant reductions in $L_{S,room}$ for R114, R115, and R116 compared to R117 (all adjusted $p \leq 0.023$); effect sizes were large ($r = -0.89$).

Figure 4 (right) presents the corresponding SNR thresholds. The untreated room again exhibited the highest median SNR requirement. Treated rooms (R115, R116) showed lower SNR thresholds

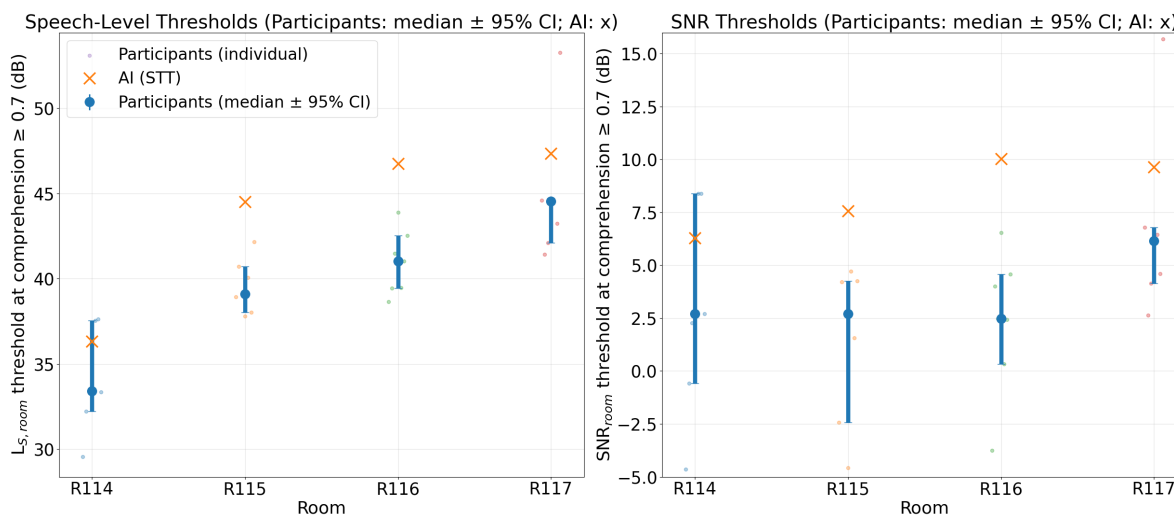


Figure 4. Speech level at the listener position corresponding to a median intelligibility ≥ 0.7 , and the corresponding SNR between reference and listener microphone.

compared to R117. Wilcoxon tests yielded large effect sizes (R115 vs. R117: $r = -0.77$; R116 vs. R117: $r = -0.70$), but differences did not remain statistically significant after Holm correction (adjusted $p = 0.141$ and $p = 0.156$, respectively). The comparison between R114 and R117 showed a smaller effect ($r = -0.32$, $p = 0.469$).

The ASR-based evaluation showed a comparable pattern for speech-level thresholds, reproducing the ordinal ranking of acoustic conditions observed in the human data. In contrast, SNR-based thresholds derived from ASR were less consistent across conditions and did not exhibit a stable ranking between treated configurations. Overall, ASR-based thresholds showed greater variability compared to the human evaluation.

Thus, while the SNR-based analysis did not demonstrate statistically robust differences after correction, the consistently large effect sizes in the treated rooms suggest a practically relevant directional trend that warrants confirmation in larger samples. Within the context of an exploratory pilot design, the consistency of effect direction across participants is considered more informative than formal statistical significance thresholds.

4. Discussion

The present study examined whether edge-based acoustic interventions modify sound field conditions in classrooms and whether such modifications translate into perceptually relevant changes beyond conventional reverberation-time-based criteria. Three principal findings emerge.

4.1. Room-Acoustic Parameters and Structural Effects

Both edge-based systems improved STI, C_{50} , and D_{50} relative to untreated conditions. These improvements were more pronounced for the absorptive system (System 1), which also produced a substantial reduction in T_{30} . The reflective/scattering system (System 2) yielded smaller changes in T_{30} but still increased early-energy-related parameters.

This pattern is consistent with previous studies showing that edge absorbers act as modal dampers and influence spatial energy distribution [11,12]. The increase in C_{50} and D_{50} indicates a relative redistribution of energy toward the early time window (< 50 ms), which is relevant for speech perception. Notably, System 2 improved clarity-related metrics despite only moderate T_{30} reduction, which is consistent with the working hypothesis that early reflection structure—not only global decay time—contributes to intelligibility.

From a physical perspective, modifications at room edges may influence boundary-related reflection patterns. Inclined or structured elements can increase angular redistribution of incident sound, while selective absorption at edges may suppress modal buildup and grazing reflections while

preserving other reflection paths. Although no direct diffuseness metrics were obtained, the observed changes in early-energy parameters are compatible with such mechanisms. This interpretation is supported by the coherence-based analysis. The ceiling-treated reference room (R114) exhibited the highest spatial coherence ($Cr = 0.31$), indicating a more structured and less diffuse late sound field. In contrast, untreated rooms showed consistently low coherence values ($Cr \approx 0.11$ – 0.12), corresponding to a highly diffuse condition. The edge-based systems modified this behavior differently: System 1 increased coherence ($Cr = 0.19$), suggesting a partial structuring of the sound field, whereas System 2 remained close to the untreated baseline ($Cr = 0.13$). The interpretation of early-energy redistribution as a diffuseness-related effect remains indirect and cannot be interpreted as a causal relationship within the present dataset.

Untreated rooms exhibited comparable STI and early-energy values, indicating similar baseline conditions. The ceiling-absorption reference (R114) represents the upper bound in conventional parameters within the investigated configurations.

4.2. Psychoacoustic Thresholds and SNR Requirements

The psychoacoustic results complement the parameter-based findings, but reveal different trends for speech level and SNR thresholds. Speech-level thresholds closely followed the behavior of T_{30} : rooms with shorter reverberation times required lower speech levels to reach the predefined intelligibility criterion, whereas untreated rooms exhibited the highest required levels.

In contrast, SNR thresholds showed a different pattern. While the untreated condition again resulted in the highest SNR requirements, the differences between treated conditions did not scale directly with reverberation time. In particular, improvements were observed even in configurations with only moderate T_{30} reduction. This divergence suggests that speech level is primarily governed by overall energy decay (i.e., reverberation time), whereas SNR requirements are influenced by additional factors related to sound field structure, particularly the temporal and spatial distribution of reflected energy.

Although SNR differences did not remain statistically significant after correction for multiple comparisons, effect sizes indicated consistent directional reductions in treated rooms. Notably, these reductions were observed even in the reflective/scattering configuration (System 2), where reverberation time remained comparatively elevated.

This pattern suggests that intelligibility under steady noise conditions is not fully captured by reverberation time alone. The coherence results indicate that these perceptual improvements do not coincide with a simple increase in diffuseness (i.e., decreasing coherence), but rather with a redistribution of sound field structure. This suggests that spatial coherence captures aspects of the acoustic field that are complementary to conventional decay-based metrics. Changes in early-energy parameters may contribute to reduced SNR requirements, although the underlying mechanisms cannot be isolated within the present dataset.

The ASR-based evaluation reproduced the ordinal ranking of acoustic conditions for speech-level thresholds; for SNR thresholds, the ranking was less stable. In contrast, the human evaluation indicated greater inter-individual variability in untreated conditions, whereas treated rooms showed more clustered threshold values, although this observation was not formally tested.

A plausible explanation for this discrepancy lies in the recording configuration. The microphone at the listening position was oriented toward the participant in order to capture the verbal response, rather than toward the sound source. Although nominally omnidirectional, the measurement microphone exhibits reduced sensitivity at higher frequencies, particularly for off-axis incidence.

As these frequency components are relevant for speech intelligibility, the recorded signal used for ASR processing does not fully represent the acoustically relevant stimulus at the listener position. This likely affected the robustness of ASR-based SNR estimation and contributed to the reduced consistency of SNR-related results compared to the human evaluation. This methodological limitation primarily affects the ASR-based analysis and should be considered when interpreting discrepancies between human and ASR-derived results.

4.3. Relation to Working Hypothesis and Implications for Classroom Acoustic Design

The working hypothesis proposed that edge-based interventions alter sound field conditions in ways not fully captured by T_{30} and that such changes may affect intelligibility under noise.

The present results are consistent with this hypothesis. Improvements in SNR thresholds were observed even in configurations where reverberation time remained comparatively elevated (System 2). These findings indicate that decay-based criteria alone may not fully describe perceptual performance in small classrooms.

From a design perspective, this suggests that edge-based modifications—whether absorptive or scattering—may represent a complementary strategy to large-area ceiling absorption. Rather than focusing solely on reducing global decay time, classroom acoustics may benefit from also considering the spatial-temporal organization of early reflections as an additional design variable.

4.4. Limitations and Future Research

The present study has several limitations. The psychoacoustic sample size was small ($N = 7$ complete datasets) and restricted to a single school environment, limiting statistical power and generalizability. Inter-individual variability, known to be substantial in child populations, could not be fully characterized.

Absolute SPL calibration of speech levels was not performed, and masking consisted of stationary broadband noise only. The field-based SNR estimation and sequential block design introduce methodological uncertainty, including potential learning or fatigue effects.

The study was conceived as an exploratory investigation to assess whether edge-based modifications plausibly influence speech intelligibility beyond conventional reverberation-time criteria under real-world conditions. The observed directional SNR effects warrant replication with larger participant cohorts across multiple schools.

Future research should include calibrated SPL measurements, randomized condition orders, and testing under fluctuating noise conditions. Most importantly, systematic investigation of structural sound field characteristics — supported by spatially resolved measurements and numerical simulations — will be required to derive a practically applicable descriptor for small-room acoustic assessment.

5. Conclusions

This field study investigated whether edge-based acoustic modifications influence speech intelligibility in classrooms beyond conventional reverberation-time-based criteria. Both edge-based systems improved STI, C_{50} , and D_{50} relative to untreated rooms. A coherence-based proxy analysis revealed systematic differences in sound field structure: untreated rooms exhibited low spatial coherence (high diffuseness), while the ceiling-treated reference showed substantially higher coherence. Edge-based systems modified coherence in a treatment-specific manner, indicating distinct structural effects on the late sound field. The absorptive system (System 1) substantially reduced T_{30} , whereas the reflective/scattering system (System 2) produced moderate T_{30} reduction while increasing early-energy-related parameters.

Psychoacoustic testing under steady noise conditions showed lower median SNR thresholds in treated rooms compared to the untreated configuration. Although SNR differences did not remain statistically significant after correction for multiple comparisons, effect sizes indicated consistent directional reductions. Improvements were observed even where reverberation time remained comparatively elevated.

These findings suggest that intelligibility under noise is not fully characterized by decay time alone, nor by diffuseness in a simple sense, but depends on the structural organization of the sound field, including the balance between early reflections and late reverberation as reflected in coherence patterns. Structural aspects of the sound field may therefore contribute to perceptual performance in small classrooms. The pilot findings provide sufficient empirical plausibility to warrant a coordinated research initiative aimed at (i) replicating the observed effects under calibrated and randomized

conditions, (ii) developing quantitative structural descriptors of early sound-field organization, and (iii) validating their perceptual relevance across diverse classroom environments. In particular, future work should aim to establish validated links between spatial coherence, perceptual thresholds, and established room-acoustic parameters, in order to determine whether coherence-based descriptors can serve as practical proxies for perceptually relevant sound field structure.

Author Contributions: The author was solely responsible for conceptualization, methodology, software development, data acquisition, formal analysis, visualization, and manuscript preparation.

Funding: This research received no external funding. The study was financed through internal resources of H2 Think gGmbH.

Institutional Review Board Statement: Ethical review and approval were waived because the study involved minimal-invasive, non-medical behavioral testing conducted in a school environment. The procedures did not exceed everyday educational activities, posed no foreseeable risk to participants, and involved no intervention or manipulation beyond standard listening tasks. No personal identifying data were collected or stored. In accordance with principles of good scientific practice, ethical considerations were assessed prior to the study, and the investigation was deemed minimal risk and compliant with applicable local regulations.

Informed Consent Statement: Informed consent was waived because the study involved minimal-invasive, low-risk activities conducted during organized holiday childcare in coordination with the responsible educational staff. Participation was voluntary, and no identifiable data were recorded. The study design did not involve medical procedures, sensitive personal data, or interventions requiring formal written consent under applicable local regulations.

Data Availability Statement: The evaluation scripts used for room impulse response processing and acoustic parameter extraction are publicly available at GitHub ([18]). The algorithm to derive the ASR-evaluation is also publicly available at GitHub ([19]) The raw psychoacoustic recordings and classroom measurement data are not publicly available due to privacy and ethical considerations but may be provided by the author upon reasonable request.

Acknowledgments: The author thanks Gerhard Ochsenfeld for establishing contact with the participating school, organizing access to the classrooms, and installing System 2 in Room 116. Michael Ochsenfeld is acknowledged for installing System 1 in Room 115. The author further thanks Gerhard Ochsenfeld for facilitating access to the rooms before and after installation for acoustic measurements. During measurement sessions, he was present on site but had no influence on study design, measurement procedures, data acquisition, analysis, or interpretation.

The author gratefully acknowledges the participating school and its educational staff for enabling access to the classrooms, coordinating measurement sessions, and supporting the implementation of the psychoacoustic testing under supervised conditions.

The author used a large language model (ChatGPT, versions 5.2 and 5.4) for language editing and formulation of parts of the manuscript. All scientific content, data analysis, and interpretation were carried out by the author. All cited sources were independently reviewed and verified.

Conflicts of Interest: The idea for this study emerged from professional discussions between the author and Gerhard Ochsenfeld, developer of System 2. Gerhard Ochsenfeld established contact with the participating school and installed System 2 in Room 116. Michael Ochsenfeld, developer of System 1, installed System 1 in Room 115. The study design was independently developed by the author. No financial compensation, payments, or material benefits were received by the author or by H2 Think gGmbH from the involved parties. The developers had no influence on data acquisition, data analysis, interpretation of results, or manuscript preparation.

References

1. Bradley, J.S. (1986). Predictors of speech intelligibility in classrooms. *Journal of the Acoustical Society of America*, 80(3), 837–845.
2. Dockrell, J.E.; Shield, B. (2006). Acoustical barriers in classrooms: The impact of noise on performance in the classroom. *Journal of the Acoustical Society of America*, 123(1), 133–144.
3. Klatt, M.; Bergström, K.; Lachmann, T. (2013). Does noise affect learning? A short review on noise effects on cognitive performance in children. *Frontiers in Psychology*, 4, 578.

4. Murgia, M.; et al. (2023). Systematic review on speech intelligibility and classroom acoustics. *Language, Speech, and Hearing Services in Schools*, 54(1), 322–335.
5. Mealings, K.; Buchholz, J. (2024). Classroom acoustics and learning: A review of recent evidence. *Trends in Hearing*, 28, 1–18.
6. International Organization for Standardization, *Acoustics – Measurement of room acoustic parameters – Part 1: Performance spaces*, ISO 3382-1:2009, Geneva, 2009.
7. DIN 18041:2016-03, *Hörsamkeit in Räumen – Anforderungen, Empfehlungen und Hinweise für die Planung*, Deutsches Institut für Normung e. V., Berlin, 2016.
8. ANSI/ASA S12.60-2010 (R2019). *Acoustical Performance Criteria, Design Requirements, and Guidelines for Schools*. Acoustical Society of America.
9. Vasconcelos Rabelo, A.T.; Santos, J.N.; Oliveira, R.C.; Magalhães, M.C. Effect of classroom acoustics on the speech intelligibility of students. *CoDAS*, 2014, 26(5), 360–366.
10. Arvidsson, T.; Brunskog, J.; Johansson, E. The effect on room acoustical parameters using a combination of absorbers and diffusers—An experimental study in a classroom. *Acoustics*, 2020, 2(3), 505–523.
11. Kraxberger, J.; Vorländer, M., *A validated finite element model for room acoustic treatments with edge absorbers*, Applied Acoustics, Vol. 208, 109390, 2023.
12. Kurz, J.; Müller-Trapet, M.; Vorländer, M., *Systematische Untersuchungen zur Funktionsweise von Kantenabsorbern als Modenbremse*, Proceedings of the DAGA 2021, Copenhagen, Denmark, 2021.
13. Yoshida, S.; Kawai, K.; Sakagami, K., *An experimental study of the performance of a crossed rib diffuser in room acoustic control*, Applied Acoustics, Vol. 181, 108145, 2021.
14. Eldien, H. H.; Hammad, R. N., *Enhancing acoustic performance of classrooms by using cymatic diffusers*, Applied Acoustics, Vol. 225, 110108, 2024.
15. Khrystoslavenko, A.; Grubliauskas, R., *Experimental studies of the sound scattering coefficient of the diffuser in the reverberation chamber*, Applied Acoustics, Vol. 207, 109337, 2023.
16. International Organization for Standardization (2004). ISO 17497-1: Acoustics – Sound-scattering properties of surfaces – Part 1. Geneva, Switzerland.
17. International Electrotechnical Commission, *Sound system equipment – Part 16: Objective rating of speech intelligibility by speech transmission index*, IEC 60268-16:2020, Geneva, 2020.
18. Kümritz, S., *Minimal Example of the Algorithms used for RIR and room acoustic parameter determination*, Available online: https://github.com/H2ThinkResearchInstitute/RIR_Evaluation (accessed on 19 February 2026).
19. Kümritz, S., *Whisper*, Available online: <https://github.com/StinkePunk/Whisper> (accessed on 19 February 2026) (accessed on 27 February 2026).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.