Sound & Vibration 2026, 60(1), 3828.
doi: https://doi.org/10.59400/sv3828

Article

Comparative acoustic analysis of treated and untreated small recording rooms using vocal and room-response measurements

Korab Shaqiri[1], Armend Xhon[2,*], Xhevdet Gashi[3], Lulkaçela Jashari[4]

[1] Faculty of Music, University of Prishtina, Prishtina 10000, Kosovo
[2] Faculty of Education, University of Prishtina, Prishtina 10000, Kosovo
[3] Faculty of Modern Music, UBT College, Prishtina 10000, Kosovo
[4] Faculty of Fashion Design, Evolution Academy, Prishtina 10000, Kosovo
*** Corresponding author:** Armend Xhoni, armend.xhoni@uni-pr.edu

CITATION

Shaqiri K, Xhoni A, Gashi X, et al. Comparative acoustic analysis of treated and untreated small recording rooms using vocal and room-response measurements. Sound & Vibration. 2026; 60(1): 3828. https://doi.org/10.59400/sv3828

ARTICLE INFO

Received: 16 December 2025
Revised: 28 January 2026
Accepted: 2 February 2026
Available online: 5 February 2026

COPYRIGHT

Copyright © 2026 Author(s). *Sound & Vibration* is published by Academic Publishing Pte. Ltd. This work is licensed under the Creative Commons Attribution (CC BY) license. https://creativecommons.org/licenses/by/4.0/

Abstract: This study investigates the influence of small-room acoustics on sung vocal quality by comparing two adjacent recording environments of nearly identical dimensions: a professionally treated room (Room A) and an untreated room (Room B). Exponential sine-sweep measurements were conducted using calibrated studio equipment, and room impulse responses were analysed using reverberation metrics (EDT, Topt, RT60), clarity indices (C50, C80), definition (D50), modal behaviour, decay characteristics, and total harmonic distortion (THD). Room A incorporates broadband absorption at primary reflection points, corner bass trapping, rear-wall diffusion, and a double-layer wall structure for low-frequency damping. In addition to room-response measurements, three sustained sung vocal phrases were recorded in both rooms at microphone distances of 10, 30, and 60 cm and analysed using spectrograms and formant trajectories. The results reveal acoustic differences between the two environments. The untreated room exhibits prolonged mid-band decay times (Topt ≈ 1.3–1.9 s), elevated EDT values, reduced clarity (C80 between −5 and +8 dB), low definition (D50 ≈ 30–40%), strong low-frequency modal buildup, and increased harmonic masking. In contrast, the treated room shows studio-appropriate decay times (Topt ≈ 0.20–0.30 s), high clarity (C80 ≈ 15–25 dB), excellent definition (D50 ≈ 90–100%), and smooth decay behaviour. Time–frequency analysis of the vocal recordings confirms that acoustic treatment stabilises temporal and spectral response, yielding cleaner harmonic structure, sharper transient articulation, improved intelligibility, and greater formant precision across all microphone distances. These findings demonstrate the critical role of acoustic treatment in achieving reliable, high-fidelity vocal recordings in small-room environments.

Keywords: room acoustics; reverberation time; clarity indices (C50, C80); sung vocal quality; spectrogram and formant analysis

1. Introduction

Recording quality in small rooms depends not only on performance, microphone technique, and production tools, but fundamentally on the acoustic behaviour of the enclosed space. In such environments, the interaction between the sound source and room boundaries generates a complex sound field commonly referred to as reverberation—a multidimensional phenomenon with physical, perceptual, and affective attributes [1]. Although widely used as an objective descriptor, reverberation correlates only partially with perceptual judgments, particularly in small rooms where

short reflection paths, strong modal behaviour, and atypical early-to-late energy ratios distinguish them from larger performance spaces [1–3]. Small rooms—such as home studios, rehearsal spaces, classrooms, or domestic environments—exhibit a highly non-diffuse sound field characterised by strong early reflections, pronounced spectral colouration, and low-frequency modal resonances [4, 5]. Human hearing is exceptionally sensitive to these variations, especially in the mid and high frequencies, where the auditory system detects subtle changes in clarity, definition, and spatial impression [5,6]. Understanding the behaviour of small rooms has become increasingly important in contemporary applications such as virtual acoustics, binaural rendering, and auditory augmented reality, where accurate modelling of early reflections and room response is essential for perceptual realism [4]. These approaches highlight that accurate representation of room acoustics requires an integrated physical–perceptual framework, since traditional metrics (e.g., RT60 alone) often fail to capture the nuanced effects of early-reflection patterns, modal decay, and nonlinear time–frequency behaviour in small volumes [2,7].

Consequently, accurate analysis of small-room acoustics requires time–frequency representations capable of capturing non-stationary behaviour and temporal smearing effects, which are not fully described by scalar reverberation metrics alone [8].

Accordingly, modern room-correction and dereverberation systems rely on precise analysis of the room impulse response (RIR), clarity metrics, and temporal decay characteristics. An accurate understanding of these parameters is essential for designing effective acoustic treatment and for predicting perceptual consequences in practical recording scenarios. In this context, the present study examines two adjacent small rooms—one professionally treated and one untreated—to quantify the effects of acoustic treatment on reverberation time, clarity indices (C50, C80), definition (D50), total harmonic distortion (THD), time–frequency behaviour, and vocal formant stability. By combining objective room measurements with detailed spectrographic analysis of sung vocal recordings captured at multiple microphone distances, this research advances the argument that only integrated physical–perceptual evaluation can adequately characterise small-room acoustics [1,2,5].

This study aims to determine the extent to which acoustic treatment transforms the temporal, spectral, and perceptual characteristics of small recording rooms, and to demonstrate how these transformations directly influence intelligibility, timbral stability, and overall fidelity in vocal recordings.

2. Literature review

Room acoustics is widely recognised as a central determinant of recording quality, influencing both the physical behaviour of sound and the perceptual experience of listeners. In small rooms—such as recording studios, rehearsal spaces, classrooms, or domestic environments—the interaction of boundary reflections, room modes, and time–frequency smearing becomes especially pronounced [9]. These environments typically exhibit non-diffuse sound fields dominated by strong early reflections, modal resonances, and uneven energy decay, all of which significantly affect timbre, spatial impression, and vocal intelligibility [5, 10]. Perceptual research

emphasises that reverberation, although traditionally treated as a physical descriptor, also comprises perceptual and affective dimensions that cannot be fully captured through conventional acoustic metrics [1, 2]. This multidimensional nature makes the study of small-room acoustics particularly complex and demands integrated physical–perceptual frameworks that consider both objective measurements and perceptual interpretation. A key distinction between small and large rooms lies in reflection patterns and modal behaviour. In small spaces, early reflections arrive with very short delays, producing pronounced comb filtering, spectral colouration, and temporal masking—effects largely absent in larger auditoria [4]. Low-frequency room modes in the 40–200 Hz range exhibit high perceptual salience due to their limited spatial distribution and long decay times. These modes interact with vocal fundamentals and formants, introducing tonal instability, harmonic distortions, and loss of spectral precision. Psychoacoustic research confirms that the auditory system interprets these anomalies not simply as variations in frequency response but as changes in spaciousness, externalisation, and emotional tone [2]. Understanding vocal behaviour in small rooms, therefore, requires analysis of both the physical artefacts produced by the environment and the perceptual mechanisms through which listeners interpret them. This duality is central to the present study.

2.1. Reverberation metrics: RT60, Topt, EDT

Reverberation time remains one of the most influential descriptors in architectural and studio acoustics. Sabine established the foundational relationship between room volume, absorption, and decay time. In contrast, later refinements such as Topt (optimal reverberation time) and EDT (Early Decay Time) provide more nuanced insights into temporal behaviour, particularly in small rooms [11]. EDT is especially relevant in such environments because it reflects the influence of early reflections on perceived intimacy and clarity. From a measurement perspective, the contemporary evaluation of reverberation relies primarily on impulse-response–based methods, following Schroeder's formulation of the energy decay curve, which enables robust estimation of decay characteristics without direct absorption measurements. In practice, achieving a full 60 dB decay is often impractical in small rooms due to background noise and non-uniform decay behaviour; consequently, alternative descriptors such as T20, T30, and especially Early Decay Time (EDT) are commonly employed. EDT, derived from the initial portion of the decay curve, provides a more stable and perceptually relevant representation of early sound-field behaviour than RT60 alone [12]. Research consistently shows that small-room reverberation is highly sensitive to both geometry and absorption distribution, leading to significant variations in decay uniformity across frequency bands. Newell recommends RT60 values of 0.15–0.30 s for professional vocal booths, noting that longer decay times introduce resonance effects that mask consonants, reduce articulation, and distort formant placement [13]. However, perceptually oriented studies caution that RT60 alone correlates only partially with subjective impressions—particularly in small rooms where early-to-late energy ratios and reflection patterns exert greater influence [1]. These findings directly inform the present study's emphasis on multifactor acoustic

analysis rather than reliance on RT60 alone.

2.2. Clarity and definition (C50, C80, D50)

Clarity metrics provide essential information on how early and late energy contribute to intelligibility and temporal precision. C50, widely used in speech analysis, reflects the balance between early (<50 ms) and late energy, with values above +5 dB indicating good intelligibility (ISO 3382-1, 2009). For musical signals, C80 is more relevant, as higher values indicate improved transient definition and separation of successive sonic events. In practical acoustic analysis, these clarity metrics are commonly evaluated across octave or one-third-octave frequency bands in order to capture their frequency-dependent behaviour, which is particularly relevant in small rooms where absorption and modal effects vary strongly with frequency [14]. Definition (D50), which expresses the proportion of early arriving energy, is another critical descriptor; values above approximately 80% are typically associated with well-treated studio environments. Bradley and Sato demonstrated strong correlations between clarity metrics and perceptual speech intelligibility, while recent work highlights their importance in environments where complex time–frequency interactions dominate [15]. Small rooms exhibit rapid fluctuations in clarity indices across listener positions, particularly when reflective surfaces or modal peaks reinforce specific frequency bands [4]. Given this sensitivity, clarity metrics are indispensable for understanding the differences between treated and untreated rooms examined in this study.

2.3. Total harmonic distortion, modal behaviour, and frequency stability

Small rooms often amplify nonlinearities in vocal and musical recordings, particularly in the low-frequency range. Beyond time-domain descriptors, frequency-domain analysis—including spectral stability and harmonic distortion—provides complementary insight into signal integrity, as distortions in the harmonic structure can significantly affect waveform fidelity and perceived sound quality [16]. Untreated environments often exhibit elevated Total Harmonic Distortion (THD) at modal frequencies (40–120 Hz), resulting in spectral peaks, unstable formants, and harmonic masking. Modal ringing not only prolongs low-frequency decay but also interacts with harmonic partials, altering timbre and producing perceptually salient artefacts. Room response equalisation research further demonstrates that uncontrolled reflections and resonances introduce colouration that compromises the intended auditory image of sound reproduction systems [7]. Since vocal signals depend heavily on stable harmonic structure for intelligibility and emotional expression, these distortions fundamentally alter the perceptual character of recorded sound. This directly motivates the inclusion of THD analysis in the present experiment.

2.4. Time–frequency behaviour, spectrograms, and formant analysis

Time–frequency visualisations such as spectrograms provide powerful tools for identifying how small rooms shape vocal behaviour. Sundberg demonstrated that formant clarity—stability, alignment, and spectral precision—depends strongly on the acoustic characteristics of the recording space [17]. Early reflections generate harmonic smearing, blurred transients, and wide-band energy dispersion, while modal

interference further contributes to spectral instability. Lokki and Pätynen found that perceptual evaluations of sound quality correlate closely with spectrographic indicators such as harmonic distinctness and smoothness of decay [18]. Contemporary work in virtual acoustics confirms that time–frequency stability is essential for maintaining perceptual authenticity, particularly in dynamic listening scenarios where room–listener interaction changes continuously [19]. Recent signal-analysis research further shows that alterations in spectrographic structure—especially in fundamental frequency and formant trajectories—can significantly influence perceptual interpretation, highlighting the analytical value of spectrogram-based evaluation beyond purely visual inspection [20]. These insights reinforce the relevance of spectrogram and formant analysis for evaluating treated and untreated environments in this study.

2.5. Objective and perceptual measures of room responses

Objective metrics—such as room impulse responses (RIRs), clarity indices, decay curves, and THD—provide essential insights into room behaviour. However, perceptually oriented studies argue that objective descriptors must be interpreted alongside listener-based assessments, especially in small rooms where traditional metrics may not fully capture subjective impressions [1]. Goetze et al. highlight the value of RIR-based quality assessment for predicting the perceptual effects of dereverberation and room-correction algorithms [21]. Similarly, augmented reality research stresses that perceptual matching of acoustic environments requires careful modelling of early reflections, diffusion patterns, and listener orientation [4]. The integration of perceptual, physical, and computational approaches, therefore, forms the contemporary foundation for understanding small-room acoustics—an approach mirrored in the methodological design of the present study.

2.6. Synthesis of the literature

The research literature consistently demonstrates that small rooms exhibit unique acoustic challenges: short reflection paths, strong modal resonances, highly variable clarity metrics, nonlinear time–frequency behaviour, and elevated harmonic distortion. These characteristics significantly affect vocal stability, intelligibility, timbral accuracy, and perceptual realism. Existing work highlights the need for integrated physical–perceptual evaluation but offers limited empirical comparisons between physically similar treated and untreated environments, particularly using sung vocal material and multiple microphone distances. The present study addresses this gap by combining room-impulse-response analysis, clarity and definition metrics, THD measurements, and detailed spectrographic examination of vocal recordings to quantify how acoustic treatment transforms the behaviour of small rooms.

3. Methodology

3.1. Experimental setup: Description of the rooms

Two test environments were selected to compare the acoustic behaviour of a professionally treated studio with that of an untreated domestic room. Both rooms have

comparable volumes and layouts, and are positioned adjacent to each other.

3.1.1. Room A—Treated professional studio

- Dimensions: 272.5 × 520 × 261 cm;
- Acoustic treatment: Broadband absorbers (walls and ceiling), Bass traps (corners), Diffusion panels (rear wall);
- Function: vocal and instrumental recording.

Acoustic treatment specification of room A

The professional studio (Room A) was treated using a combination of broadband absorption, low-frequency trapping, controlled diffusion, and structural low-frequency damping, aiming to achieve a balanced decay profile, a stable early-reflection field, and reduced low-frequency modal excitation. Broadband absorbers were installed on the primary reflection surfaces of the side walls and ceiling. These absorbers consisted of porous materials (mineral wool / fibreglass-based panels) with typical thicknesses up to approximately 100 mm, designed to provide adequate sound absorption across the mid- and high-frequency range (≈200 Hz–10 kHz). The porous structure of the absorbers increases viscous losses in air-particle motion, resulting in high absorption coefficients in the critical speech and vocal frequency bands.

Low-frequency control in Room A was achieved through a combination of corner bass trapping and structural wall design. Bass traps were installed in the room corners, where sound pressure maxima of axial and tangential modes are expected. These corner traps function as broadband low-frequency absorbers, extending absorption effectiveness into the 50–200 Hz range and thereby reducing modal ringing, peak buildup, and prolonged decay times in the bass region. In addition, the room features double-layer gypsum board construction with high-density Knauf Diamond plasterboards, resulting in an overall wall thickness of approximately 20 cm. This double-leaf wall system introduces mass–spring–mass behaviour, providing additional damping of low and sub-bass frequencies. Such constructions are particularly effective at attenuating frequencies below approximately 100 Hz, helping control sub-frequency energy that is difficult to address with porous absorbers alone. The structural damping provided by the wall system, therefore, complements the performance of the bass traps by further reducing low-frequency resonance and stabilising decay behaviour in the sub-bass region.

Diffusion panels were installed on the rear wall of the room. The purpose of diffusion was not to increase absorption, but to redistribute reflected energy spatially and temporally, reducing strong specular reflections and improving the uniformity of the sound field. This configuration contributes to a more even energy decay and helps preserve natural spatial characteristics without introducing excessive reverberation. The combined effect of broadband absorption, bass trapping, structural low-frequency damping, and controlled diffusion results in an increased equivalent absorption area (A), a higher average absorption coefficient ($\bar{\alpha}$), and a more uniform decay behaviour across frequency bands. These treatment strategies are consistent with established design principles for small professional recording rooms and vocal booths, particularly in applications requiring accurate vocal and speech recording.

3.1.2. Room B—Untreated room

- Dimensions: 280 × 540 × 328 cm;
- No acoustic treatment or isolation elements;
- Similar floor plan to Room A, used as the reference untreated environment.

Equipment

All measurements and recordings were performed using identical signal chains in both rooms:

- Microphone: Neumann U87 Ai;
- Mic Preamp: API 512c (API 500V rack);
- Audio Interface: Apogee Symphony I/O 16 × 16;
- Monitoring Loudspeaker: Genelec 8351;
- Measurement Microphone: dbx RTA-M (calibrated);
- Cabling: VOVOX Sonorus XLR;
- Software: REW v5.31.3;
- Computer: Apple Studio M1 Max;
- Microphone Height: 1.15 m;
- Speaker–Microphone Distance: 2.0 m.

The microphone height, loudspeaker position, and source–receiver geometry were kept strictly identical in both rooms to ensure experimental comparability. Previous research on the reproducibility of voice parameters has demonstrated that room acoustics and background noise can exert a more substantial influence on measured voice outcomes than microphone type, provided that the recording chain is kept constant [22]. Accordingly, the use of an identical calibrated microphone, fixed placement geometry, and a consistent signal chain in the present study ensures that the observed differences between Room A and Room B are attributable primarily to room-acoustic conditions, with recording hardware–related variability minimised through controlled experimental design.

3.2. Measurement procedure

Impulse response (IR) measurements

Impulse responses were captured in both rooms using a full-range sine sweep from 20 Hz to 20 kHz. Each impulse response measurement was repeated three times in both rooms, and the resulting responses were averaged to minimise random noise and ensure measurement consistency. Although the excitation signal covered the full audible spectrum (20 Hz–20 kHz), reverberation metrics were evaluated in accordance with ISO 3382 conventions, with primary interpretation focused on the 125 Hz–4 kHz octave bands. From these IRs, the following parameters were derived:

- RT60, Topt, and EDT;
- C50 and C80 (clarity indices);
- D50 (definition);
- Energy-Time Curve (ETC);
- Energy Decay Curve (EDC, derived using the Schroeder backward integration);
- Time of peak energy arrival;

- Decay curvature and decay-slope deviation metrics.

THD (total harmonic distortion)

Harmonic distortion (H2–H9) was measured using swept sine responses recorded through the Genelec 8351 monitor. This allowed comparison of:

- Non-linearities excited by the room;
- Harmonic masking;
- Low-frequency distortion amplification in the untreated room.

3.3. Vocal recording procedure

A single vocalist performed the identical vocal phrase at three fixed distances:

- 10 cm;
- 30 cm;
- 60 cm.

Recordings were captured in both rooms using identical gain staging. Each vocal phrase was recorded three times at each microphone distance, and the most stable take (in terms of pitch and amplitude consistency) was selected for analysis. These recordings were analysed for:

- Spectral balance;
- Formant stability and harmonic visibility;
- Temporal and spectral smearing;
- Energy decay patterns in the spectrogram;
- Envelope consistency across time;
- Impact of reverberation on clarity.

3.4. Vocal material

The vocal material used in this experiment consisted of a short excerpt from the Albanian art song "Rrjedh në këngë e ligjërime" ("Flows in Songs and Discourses"). The piece, composed by Feim Ibrahimi with lyrics by Gjok Beci, holds a distinguished place in Albanian musical heritage [23]. Originally performed by Vaçe Zela and later reinterpreted by notable Kosovar singers such as Ilir Shaqiri and Shkurte Fejza, the song is widely recognised for its expressive melodic line and culturally significant poetic content. This song was selected for the acoustic experiment for several reasons:

1. Melodic suitability: The phrase includes sustained vowels (/a/, /e/, /i/, /o/) and articulated consonants, enabling clear observation of both formant structure and articulatory transients. Sustained vowels provide stable spectral and phonatory conditions for formant-based acoustic analysis [24], while articulated consonants—particularly fricatives—are sensitive to fine articulatory precision and transient spectral behaviour [25];
2. Dynamic and spectral balance: The melody provides both soft and moderately intense segments, which are valuable for analysing spectral decay and the influence of reverberation on vocal timbre;
3. Cultural neutrality in vibrato and phrasing: The selected excerpt allows the vocalist to maintain stable pitch and controlled vibrato, conditions that improve

accuracy in spectrogram and energy decay analysis;

4. Poetic clarity: The text of the song features open vowel structures and minimal prosodic complexity, making it ideal for investigating clarity indices (C50, C80), D50, and intelligibility metrics.

Artistically, the song stands as a symbolic representation of Albanian cultural memory and national identity. Its lyrical narrative reflects themes of continuity, guardianship of tradition, and emotional resilience. From an engineering perspective, its balanced phrase structure, well-formed formants, and clear articulation make it an excellent choice for controlled acoustic comparison between treated and untreated rooms.

3.5. Parameters analysed

The study includes a multi-dimensional acoustic analysis covering:

Room acoustics

- RT60, Topt, EDT;
- C50, C80 clarity indices;
- D50 definition;
- Centre Time (Ts);
- Energy Decay Curves;
- Schroeder Integral;
- ETC and peak reflections.

Signal distortion

- THD (H2–H9);
- Masking effects;
- Nonlinear behaviour under reverberation load.

Time–frequency behaviour

- Spectrograms (treated vs untreated);
- Energy distribution;
- Formant visibility;
- Harmonic decay comparison;
- Peak-energy mapping.

Microphone-to-source behaviour

- Effect of room reflections at 10/30/60 cm;
- Vocal clarity changes;
- Changes in dynamic stability.

4. Results

4.1. Reverberation time and decay characteristics

4.1.1. Results of room B—Untreated environment

Measurements conducted according to ISO 3382-1 [26] reveal that the untreated room exhibits prolonged and unstable reverberation:

- Topt varies between 1.30–1.90 s, substantially above recommended values for critical listening and vocal recording environments [27];
- EDT ≈ 0.65 s, demonstrating a slow early decay, typically associated with poor articulation [28];
- The longest decays occur in the 60–120 Hz and 200–300 Hz ranges, corresponding with strong room modes predicted for small rectangular rooms [10];
- Schroeder decay curves show non-linear slopes, multiple decay rate inflexions, and signs of modal dominance (**Figure 1**);
- Time–frequency analysis indicates persistent modal ringing, reduced energy dissipation, and uneven decay across the spectrum (**Figure 2**).

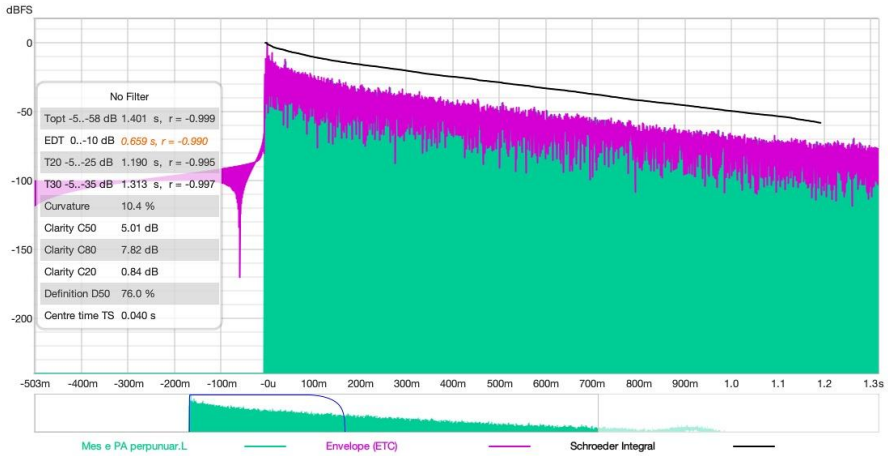

Figure 1. Impulse Response in untreated Room B showing prolonged decay and modal interference.

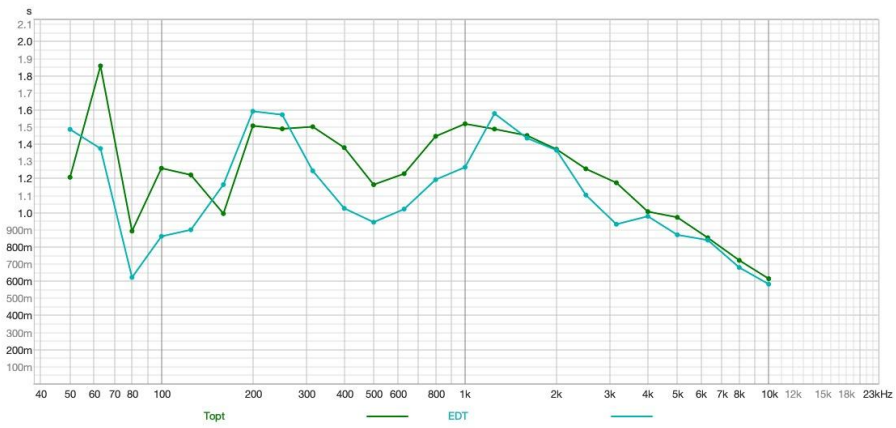

Figure 2. RT60 and EDT curves in untreated Room B showing long and uneven decay behaviour.

These characteristics confirm the expected behaviour of non-treated domestic-sized spaces: excessive reverberation, uncontrolled modal buildup, and poor temporal decay uniformity.

4.1.2. Results of room A—Treated professional studio

The treated room displays a decay behaviour consistent with professionally optimised rooms:

- Topt between 0.20–0.30 s, suitable for close-mic vocal recording and precise

monitoring [29];

- EDT ≈ 0.14–0.20 s, reflecting a controlled early reflection field and efficient absorption;
- Schroeder curves show smooth, linear decay without modal irregularities (**Figure 3**);
- Low-frequency resonances are strongly suppressed, and no modal ringing is visible in spectrogram analysis (**Figure 4**).

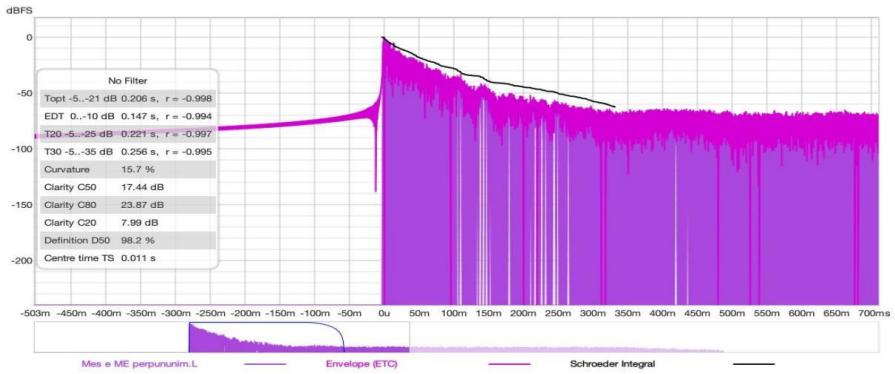

Figure 3. Impulse Response in treated Room A showing rapid, controlled decay.

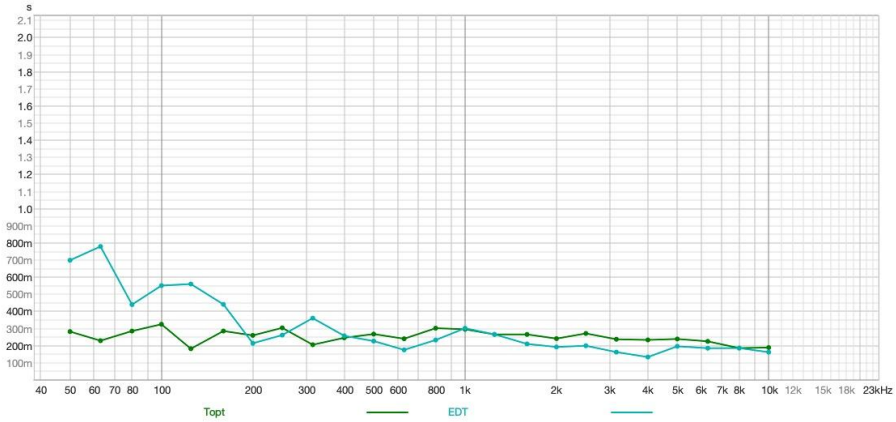

Figure 4. RT60 and EDT curves in treated Room A showing rapid and controlled decay.

These results demonstrate that Room A provides a neutral, controlled, and temporally stable decay field, meeting international acoustic guidelines for small recording studios.

4.2. Frequency-band-specific acoustic behaviour

To better understand how acoustic treatment influences room performance, the measured parameters were analysed across three frequency regions: low (50–200 Hz), mid (200 Hz–2 kHz), and high (2–20 kHz). This frequency-band-oriented analysis provides insight into the mechanisms through which broadband absorption, bass trapping, and diffusion affect reverberation behaviour, clarity metrics, and harmonic stability.

4.2.1. Low-frequency range (50–200 Hz)

In the untreated room (Room B), the low-frequency band exhibits the most pronounced acoustic degradation. RT60 and Topt values are significantly elevated

in this range, with extended decay times concentrated around the dominant axial and tangential modes within the measured range (≈50–80 Hz). These prolonged decays are accompanied by strong modal ringing, uneven spatial energy distribution, and elevated Total Harmonic Distortion (THD), particularly between 50 and 120 Hz. Such behaviour reflects insufficient low-frequency absorption and is characteristic of lightly damped rectangular enclosures. In contrast, the treated room (Room A) demonstrates a substantially improved low-frequency response. Installing corner bass traps increases modal damping, reducing the Q-factor of dominant room modes and shortening low-frequency decay times. Although modal frequencies themselves remain unchanged due to room geometry, their decay behaviour becomes more controlled, as evidenced by smoother RT60 and EDT curves and reduced low-frequency THD. This improvement directly contributes to reduced harmonic masking and increased stability of low-order vocal harmonics.

Within the measured low-frequency band, the improvement observed below approximately 100 Hz in Room A is supported by both corner bass trapping and the structural low-frequency damping introduced by the double-layer wall construction.

4.2.2. Mid-frequency range (200 Hz–2 kHz)

The mid-frequency band, which encompasses the most perceptually critical region for vocal intelligibility and formant definition, shows marked differences between the two environments. In Room B, broadband reflections from untreated surfaces lead to increased early reflection interference, resulting in reduced clarity metrics (C50, C80) and low definition values (D50). These effects manifest as temporal smearing and instability in formant trajectories, particularly during vowel transitions. Room A exhibits significantly improved performance in this range. Broadband absorbers installed on primary reflection points effectively attenuate early reflections, increasing the direct-to-reverberant energy ratio. As a result, clarity indices (C50 and C80) rise substantially, and D50 values exceed recommended thresholds for professional vocal recording. The controlled mid-frequency decay directly enhances articulation, intelligibility, and spectral precision in the recorded vocal signal.

4.2.3. High-frequency range (2–20 kHz)

At high frequencies, untreated surfaces in Room B generate strong specular reflections that contribute to comb filtering and transient blurring. Although decay times in this range are shorter than in the low band, uncontrolled reflections still impair temporal precision and reduce the perceived sharpness of vocal onsets. In Room A, the combination of broadband absorption and rear-wall diffusion leads to a more uniform high-frequency energy distribution. Absorption reduces excessive brightness and flutter echoes, while diffusion redistributes residual reflected energy without introducing additional reverberation. This balance preserves transient clarity and supports natural spatial impression, resulting in clean attack transients and consistent high-frequency behaviour across repeated measurements.

4.2.4. Summary of frequency-band effects

Overall, acoustic treatment improves room performance through complementary mechanisms acting across different frequency regions. Bass trapping primarily

addresses low-frequency modal decay and harmonic distortion, broadband absorbers enhance mid-frequency clarity and intelligibility, and diffusion supports high-frequency spatial uniformity. This frequency-specific interaction explains why Room A exhibits uniformly superior acoustic metrics and more stable vocal recordings compared to the untreated environment.

4.3. Clarity and definition metrics

4.3.1. Room B—Untreated

Clarity indices reveal severe degradation in the intelligibility and transparency of the recorded signal:

- C50 ≈ 5 dB, well below the +10 dB threshold for speech intelligibility [26];
- C80 between 5 and +8 dB, indicating inconsistent clarity across frequency bands;
- D50 ≈ 30–40%, significantly lower than the recommended minimum of 80% for vocal recording [27].

Such values correspond to blurred articulation, smeared transients, and poor consonant definition, making the room unsuitable for accurate recording (**Figure 5**).

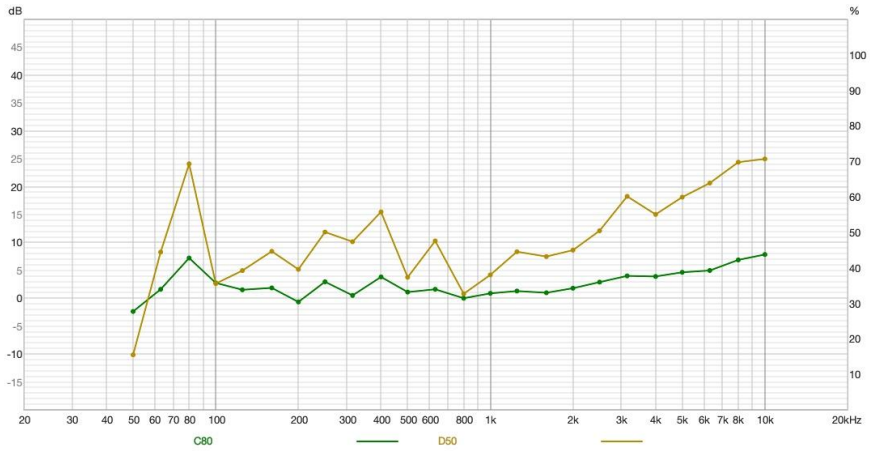

Figure 5. C80 and D50 values in untreated Room B indicate low clarity.

4.3.2. Room A—Treated

Room A exhibits excellent clarity performance:

- C50 ≈ 17 dB, providing strong support for vocal intelligibility;
- C80 between 15–25 dB, consistent across all measured bands;
- D50 ≈ 90–100%, indicative of an extremely favourable direct-to-reverberant energy ratio.

High clarity and definition reflect accurate reproduction of transient content, essential for natural and intelligible vocal recordings [29] (**Figure 6**).

4.4. Total harmonic distortion (THD)

THD measurements highlight a direct correlation between room modes and harmonic distortion:

- Room B shows elevated THD in the 40–120 Hz region, aligned with modal resonances (**Figure 7**). Such behaviour distorts the spectral balance and masks

low-order harmonics [28];

- Room A displays smoother low-frequency THD curves and no significant distortion peaks (**Figure 8**).

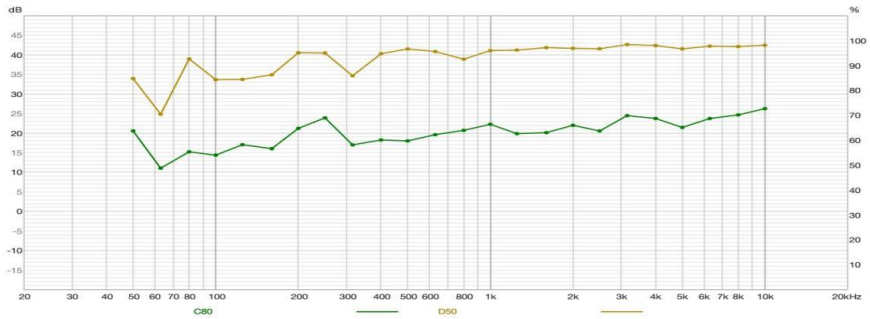

Figure 6. C80 and D50 values in treated Room A indicate high articulation.

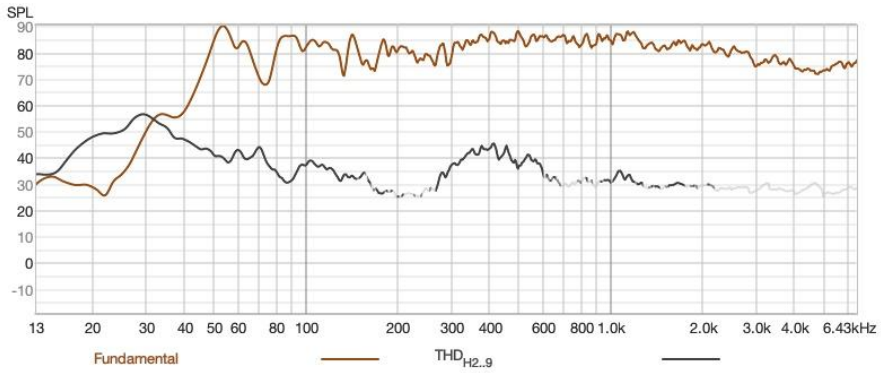

Figure 7. THD distribution in untreated Room B with elevated low-frequency distortion.

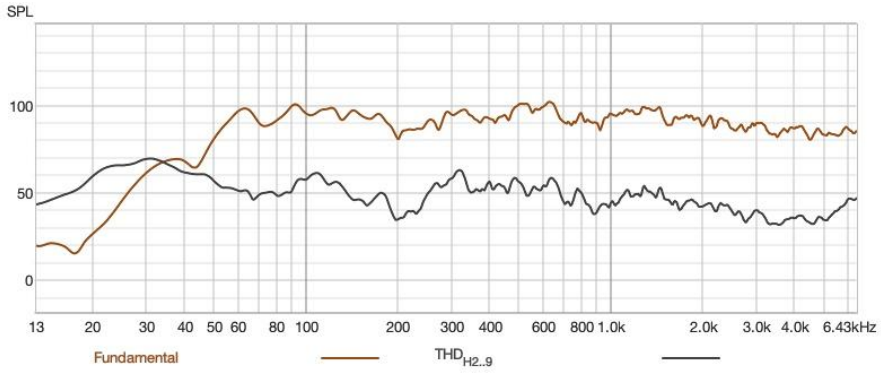

Figure 8. THD distribution in treated Room A with stable harmonic behaviour.

The treated environment, therefore, ensures clearer harmonic representation, which is essential for precise vocal timbre reproduction.

4.5. Spectrogram and time–frequency behaviour

4.5.1. Analysis of room B—Untreated

Spectrogram analysis reveals several detrimental phenomena:

- Strong harmonic smearing across time;
- Formant instability, especially in vowel transitions;
- Energy dispersion causing loss of phonetic detail;

- Persistent low-frequency buildup;
- Blurred transient onsets, indicative of a mechanically reflective space.

These effects align with known issues of early reflection interference and modal colouration in untreated rooms [5] (**Figure 9**).

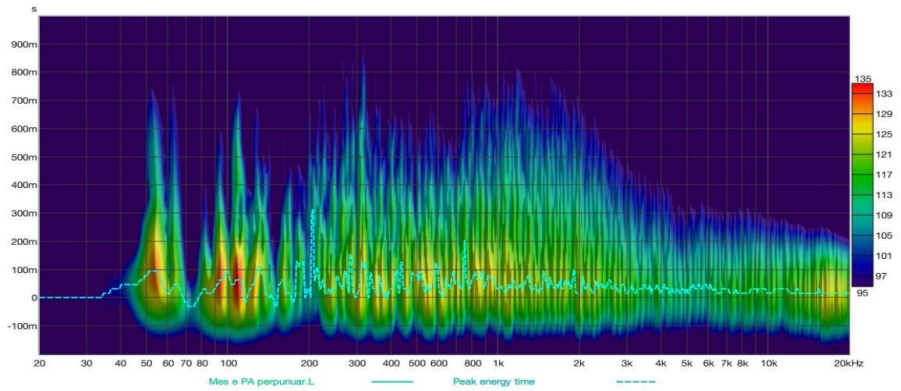

Figure 9. Room B spectrogram showing harmonic smearing and modal buildup.

4.5.2. Analysis of room A—Treated

In contrast, Room A presents:

- Clearly resolved harmonic series;
- Stable formant structure;
- Very clean and sharply defined attack transients;
- Absence of modal blurring;
- High consistency across repeated vocal takes.

The treated room maintains time–frequency integrity, a defining characteristic of high-quality recording environments (**Figure 10**). Additional room-acoustic measurements and extended spectrogram visualisations are provided in **Appendix A**.

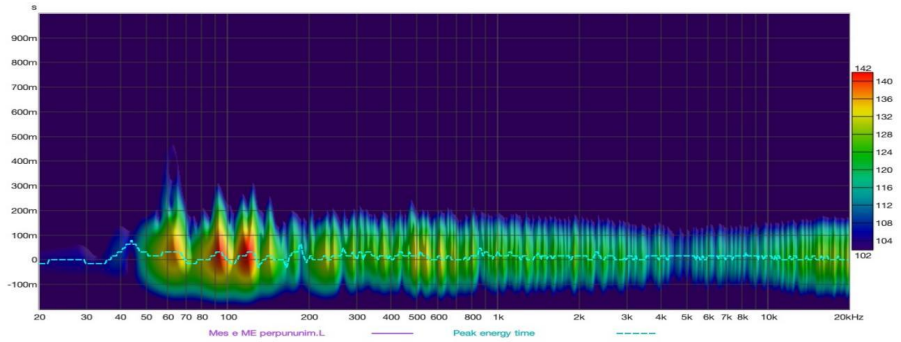

Figure 10. Room A spectrogram showing clean transients.

4.6. Vocal distance analysis

This section examines the influence of microphone distance on sung vocal quality across both recording environments. Supplementary vocal signal analyses and high-resolution spectrograms corresponding to these recordings are presented in **Appendix B**.

4.6.1. 10 cm distance

- Room A: highest articulation, negligible colouration, excellent proximity consistency;
- Room B: strong comb filtering and early-reflection interference, easily visible in spectrograms.

4.6.2. 30 cm distance

- Room A: stable spectral balance with mild room contribution;
- Room B: increased colouration, spectral imbalance, and higher reverberant energy.

4.6.3. 60 cm distance

- Room A: remains usable, natural room tone without masking;
- Room B: reverberation overwhelms the signal; formants become masked; intelligibility drops dramatically.

These findings align with established proximity and room-gain interactions documented in vocal acoustics research [17].

5. Discussion

The experimental results show a strong alignment between the measured behaviour of the two rooms and classical models of small-room acoustics. In accordance with modal theory [5, 9], the untreated Room B behaves as a lightly damped rectangular enclosure with significant low-frequency modal buildup, extended reverberation, and non-diffuse decay. In contrast, the treated Room A exhibits a highly damped, near-critically controlled acoustic field.

5.1. Relationship between acoustic treatment materials and measured acoustic parameters

The measured reverberation and clarity characteristics of Room A are directly related to the physical properties and placement of the acoustic treatment materials used. In small rooms, reverberation behaviour is strongly governed by the equivalent absorption area (A) and the spatial distribution of absorption, rather than room volume alone. Introducing broadband absorbers, bass traps, and diffusion in Room A substantially alters these parameters, producing a controlled and predictable acoustic response. Broadband absorbers installed on primary reflection surfaces increase sound absorption predominantly in the mid- and high-frequency bands. By attenuating early reflections arriving within the first 50–80 ms, these absorbers reduce early reflection interference and increase the direct-to-reverberant energy ratio. This mechanism directly explains the observed reduction in EDT and the substantial improvement in clarity metrics (C50 and C80) measured in Room A. Lower EDT values indicate that early energy decays rapidly, which perceptually translates into improved articulation, intelligibility, and transient precision.

The presence of corner bass traps primarily influences low-frequency acoustic behaviour. In untreated Room B, insufficient low-frequency absorption allows axial and tangential room modes to persist, resulting in extended low-frequency decay times and elevated RT60/Topt values. In Room A, bass traps increase low-frequency

damping by introducing porous absorption at pressure maxima, thereby reducing modal Q-factors and shortening decay times below 200 Hz. This effect is clearly reflected in smoother low-frequency decay curves, reduced modal ringing, and lower Total Harmonic Distortion values. In addition to porous absorption and corner bass trapping, the double-layer gypsum board wall construction contributes to low-frequency damping through mass–spring–mass behaviour, which is particularly effective below approximately 100 Hz. This structural damping mechanism helps explain the smoother low-frequency decay and reduced sub-bass resonance observed in the treated room, complementing the performance of conventional bass traps.

Diffusion panels installed on the rear wall contribute to acoustic control through a different mechanism. Rather than reducing total energy, diffusion redistributes reflected sound both spatially and temporally, minimising strong specular reflections and improving sound-field uniformity. This redistribution supports stable clarity metrics across listener and microphone positions and helps preserve natural spatial characteristics without increasing reverberation time. In combination with absorption, diffusion therefore plays a stabilising role in the temporal structure of the impulse response and contributes to the uniform decay observed in Room A.

The combined action of these treatment elements results in a substantial increase in the average absorption coefficient ($\bar{\alpha}$) and equivalent absorption area (A), as reflected in the short RT60 and Topt values measured in Room A. Importantly, the improvement is not uniform across frequency bands but arises from frequency-dependent mechanisms: bass traps dominate low-frequency control, broadband absorbers govern mid-frequency clarity and intelligibility, and diffusion stabilizes high-frequency spatial behaviour. This multi-layered interaction explains why Room A achieves consistently superior acoustic performance across all evaluated parameters, whereas the untreated room remains dominated by modal and reflective artefacts.

Reverberation time modelling: Sabine and Eyring

The classical Sabine formula can approximate the reverberation time RT60 for a room:

$$\mathcal{RT}_{60,Sabine} = 0.161 \cdot \frac{\mathcal{V}}{\mathcal{A}}$$

where:

$\mathcal{V}$ is the room volume (m^3) and $\mathcal{A} = \sum_i \boldsymbol{a_i}\, \boldsymbol{S_i}$ is the equivalent absorption area (in sabins), with $\boldsymbol{a_i}$ being the absorption coefficient of surface $\boldsymbol{S_i}$.

For higher overall absorption, the Eyring formula provides a more accurate estimate [30]:

$$\mathcal{RT}_{60,Eyring} = 0.161 \cdot \frac{\mathcal{V}}{-\mathcal{S} ln(1-\bar{\alpha})}$$

where:

$\mathcal{S}$ is the total surface area of the room and

$\bar{\alpha} = \boldsymbol{A/S}$ is the average absorption coefficient.

For the two rooms in this study, the approximate geometrical parameters are:

Room A (treated):

$L_x \approx 2.73$ m, $L_y \approx 5.20$ m, $L_z \approx 2.61$ m

$V_A \approx 37\ m^3$, $S_a \approx 70\ m^2$

Room B (untreated):

$L_x \approx 2.80$m, $L_y \approx 5.40$ m, $L_z \approx 3.28$ m

$V_B \approx 37 m^3$, $S_B \approx 84\ m^2$

Using the measured RT60/Topt values, the Sabine and Eyring relationships can be inverted to estimate the effective absorption area of the room, as commonly applied in room-acoustic analysis [5,30]:

$$A \approx 0.161 \cdot \frac{V}{RT_{60}} \Rightarrow \bar{\alpha} = \frac{A}{S}$$

For Room B, the long measured reverberation times (Topt in the order of ~1 s and above in the mid band) imply a small A and low $\bar{\alpha}$, characteristic of hard, reflective surfaces and minimal absorption. In contrast, the short RT60 in Room A (≈0.2–0.3 s in the mid-range) implies a much larger equivalent absorption area and hence a significantly higher average absorption coefficient. In practical terms, the treatment in Room A increases $\boldsymbol{A}$ by several times compared to Room B, consistent with professional studio designs [31].

Eyring's formulation is particularly suitable for the treated room, where $\bar{\alpha}$ is relatively high, and Sabine's linear approximation becomes less accurate. The close correspondence between measured and theoretically predicted RT60 in Room A confirms that the installed acoustic treatment behaves as an effective distributed absorber across the mid- and high-frequency bands.

5.2. Equivalent absorption and treatment efficiency

By comparing the relative RT60 values between the two rooms, the ratio of effective absorption areas can be qualitatively inferred:

$$\frac{A_A}{B_B} \approx \frac{RT_{60,B}}{RT_{60,A}}$$

Given that Room B exhibits reverberation times several times longer than those of Room A in the midfrequency range, the treated environment effectively provides multiple times the absorption area of the untreated one. This is fully consistent with expectations for a professionally treated studio employing broadband absorbers, bass traps, and controlled reflection points. This increase in absorption not only reduces RT60 but also smooths the decay curve, yielding a more linear and frequency-uniform energy decay. The measured Schroeder curves in Room A exhibit almost exponential decay, whereas those in Room B show "broken" decay lines and modal plateaus, confirming the non-diffuse, under-damped character of the untreated room.

5.3. Modal frequency analysis

The axial modal frequencies of a rectangular room can be computed using:

$$f_{n_x\ n_y\ n_z} = \frac{c}{2}\sqrt{\left(\frac{n_x}{L_x}\right)^2 + \left(\frac{n_y}{L_y}\right)^2 + \left(\frac{n_z}{L_z}\right)^2}$$

where

c is the speed of sound (≈343 m/s),

L_x, L_y, L_z are the room dimensions, and

n_x, n_y, $n_z \in \mathrm{N}_0$ are the mode indices.

Table 1 shows the first axial modes (n = 1 along each axis, others zero) for the two rooms:

Table 1. First axial modal frequencies (Hz) for Room A and Room B.

Axis/mode	Room A (treated)	Room B (untreated)
f_{100} *(x-axis)*	≈63 Hz	≈61 Hz
f_{010} *(y-axis)*	≈33 Hz	≈32 Hz
f_{001} *(z-axis)*	≈66 Hz	≈52 Hz

These modes fall squarely within the bass range (30–80 Hz), where most small-room problems occur [5, 9]. In Room B, the lack of low-frequency absorption and trapping allows these modes to be strongly excited, leading to the following:

- extended modal ringing in the low band;
- uneven spatial distribution of bass energy;
- prominent peaks and dips in the frequency response;
- coupling between room modes and vocal fundamentals or low harmonics.

The vertical mode in Room B (≈52 Hz) is particularly problematic, as it interacts with common male vocal fundamentals and low musical content, causing audible "boomy" colouration and long decay visible in the energy-decay and spectrogram plots. In contrast, the treatment in Room A—especially any bass trapping and broadband absorption—effectively increases the damping of these modes. Although the modal frequencies themselves are determined purely by geometry and thus remain the same, their Q-factor and decay time are reduced, which is reflected in:

- smoother low-frequency decay in Topt/RT60 curves;
- absence of strong low-frequency "tails" in the waterfall and spectrograms;
- more uniform low-frequency energy distribution across time.

5.4. Integration with measured vocal behaviour

The theoretical RT and modal analyses are fully consistent with the measured vocal data. In Room B, the combination of long RT60, low clarity indices (C50, C80), and strong low-frequency modes causes:

- smeared transient response and reduced time-resolution of consonants;
- formant instability due to interference between direct and reflected sound [17];
- harmonic masking and colouration visible as blurred, vertically "thickened" harmonics in the spectrograms;
- reduced Definition (D50), in line with Bradley and Sato's findings on the relationship between clarity metrics and intelligibility [18].

In Room A, the short RT60, high clarity (C50/C80), and high D50 are exactly what would be expected for a room whose effective absorption area meets or exceeds design guidelines for professional vocal booths [26]. Spectrograms show:

- clean, horizontally well-defined harmonics;
- stable formant trajectories without significant spectral smearing;
- sharply defined transient onsets;
- minimal low-frequency ringing.

These observations align with perceptual models and listening-based studies showing that environments with controlled early reflections and uniform decay yield higher ratings of clarity, naturalness, and timbral fidelity [18]. In engineering terms, acoustic treatment in Room A does not merely "improve the room" in a qualitative sense; it reconfigures the underlying acoustic field by:

- reducing effective RT60 and EDT across bands;
- increasing equivalent absorption area and average absorption coefficient;
- heavily damping problematic low-frequency modes;
- stabilising the time–frequency representation of the vocal signal.

As a result, the recorded voice in Room A is a much more accurate representation of the source, while the recording in Room B is significantly dominated by room-induced artefacts.

The findings of this study carry significant practical value for audio engineering, music education, vocal performance, and studio design. The clear acoustic differences identified between treated and untreated small rooms demonstrate that even minimal but well-designed treatment can substantially improve recording accuracy, speech intelligibility, and vocal timbral stability. For practitioners involved in home-studio construction, the results highlight the importance of early-reflection control and broadband low-frequency absorption as the primary interventions that yield immediate perceptual benefits. In music education and vocal coaching, the study reinforces the need for acoustically controlled environments to ensure accurate vocal monitoring, formant training, and reliable assessment of tone production. Similarly, broadcasters and content creators can use these insights to optimise recording spaces for clarity and consistency, reducing post-processing time and improving overall audio quality. Ultimately, the research underscores that acoustic treatment is not merely a technical upgrade but an essential component for achieving professional, accurate, and reproducible sound in any small recording environment.

6. Conclusion

The analysis demonstrates that professional acoustic treatment fundamentally transforms the behaviour of small recording rooms. Reverberation times are reduced by more than 80%, producing a controlled and predictable acoustic field that enhances both speech and music recording. Objective clarity metrics (C50, C80) and definition (D50) improve substantially, reflecting major gains in intelligibility, transient accuracy, and temporal precision. Spectrograms and energy decay curves confirm a significant reduction of smearing, modal ringing, and low-frequency buildup in the treated environment. These improvements translate directly into higher vocal recording quality, where formant stability, spectral fidelity, and articulation are markedly superior in Room A. Moreover, the study shows that microphone distance interacts strongly

with room acoustics: untreated rooms degrade rapidly beyond 10–30 cm, whereas treated rooms maintain stability even at 60 cm. Harmonic behaviour and THD patterns are also more consistent in the treated space, preserving timbre integrity. These observations are consistent with recent voice-acoustics research demonstrating that increased reverberation destabilises harmonic structure and degrades objective voice quality measures, while low-reverberation environments preserve spectral integrity and formant-related stability [32].

Beyond improvements in individual acoustic metrics, the findings highlight the decisive roles of acoustic treatment materials and room geometry in shaping the functional performance of small rooms. The results demonstrate that acoustic treatment acts through frequency-dependent mechanisms rather than as a uniform adjustment. Broadband absorbers primarily improve mid-frequency behaviour by attenuating early reflections and reducing early decay time, which directly enhances clarity indices and vocal intelligibility. In parallel, low-frequency bass trapping mitigates modal dominance by reducing modal Q-factors and shortening decay times below 200 Hz, thereby stabilising harmonic behaviour and limiting low-frequency masking. This frequency-dependent action of acoustic treatment aligns with clinical and experimental findings showing that room acoustic parameters—particularly reverberation time and clarity—exert a decisive influence on the reliability and stability of objective voice quality metrics [33].

Room geometry further conditions the effectiveness of these treatments. Although modal frequencies are determined by room dimensions and remain unchanged, their perceptual impact is strongly influenced by the spatial distribution and effectiveness of absorption. The treated room illustrates how appropriate placement of absorptive materials at early reflection points and pressure maxima can significantly overcome the inherent acoustic limitations of small rectangular spaces. Diffusion complements absorption by redistributing reflected energy without increasing reverberation time, contributing to spatial uniformity and preserving natural acoustic impressions.

Overall, the study confirms that optimal acoustic performance in small recording rooms emerges from the integrated design of material properties, spatial placement, and room geometry. Acoustic treatment is therefore not an optional enhancement, but a fundamental structural requirement for achieving reliable, reproducible, and professional-grade vocal and instrumental recordings in small-room environments.

6.1. Limitations

Although the findings of this study are robust, several limitations must be acknowledged. The experiment relied on a single vocalist, a single monitoring loudspeaker, and two rooms of similar size; therefore, the results may not generalise to larger spaces, multiple performers, or different loudspeaker configurations. Additionally, the analysis focused on steady vocal phrases rather than full musical passages, which may exhibit different temporal, spectral, and dynamic behaviours. Nevertheless, the controlled experimental design was essential for isolating room-acoustic effects with high precision.

6.2. Future work

Future research could expand this investigation by incorporating multiple vocalists, microphone types, and musical instruments, as well as analysing larger and more varied room geometries. Binaural and psychoacoustic listening tests would provide deeper insight into subjective impressions of clarity, spatial perception, and timbral accuracy. Further studies may also explore advanced computational room-correction methods and the interaction between digital processing and physical acoustic treatment, enabling more comprehensive models of small-room recording environments.

Author contributions: Conceptualization, AX and KS; methodology, AX and KS; experimental measurements, XG and KS; acoustic design and material implementation, LJ; data analysis, AX and KS; visualization, KS and LJ; writing—original draft preparation, AX; writing—review and editing, AX, KS, XG, and LJ; supervision, AX. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional review board statement: Not applicable.

Informed consent statement: Not applicable.

Data availability statement: The data presented in this study are available from the corresponding author upon reasonable request.

Conflict of interest: The authors declare no conflict of interest.

References

1. Kaplanis N, Bech S, Jensen S, et al. Perception of reverberation in small rooms: A literature study. In: Proceedings of the Audio Engineering Society (AES) 55th International Conference; 27–29 August 2014; Helsinki, Finland.
2. Zahorik P. Spatial Hearing in Rooms and Effects of Reverberation. In: Litovsky RY, Goupell MJ, Fay RR, et al. (editors). Binaural Hearing, Springer Handbook of Auditory Research. Springer International Publishing; 2021. pp. 243–280. doi: 10.1007/978-3-030-57100-9_9
3. Barron M. Auditorium Acoustics and Architectural Design, 2nd ed. Spon Press; 2024.
4. Neidhardt A, Schneiderwind C, Klein F. Perceptual Matching of Room Acoustics for Auditory Augmented Reality in Small Rooms—Literature Review and Theoretical Framework. Trends in Hearing. 2022; 26: 23312165221092919. doi: 10.1177/23312165221092919
5. Kuttruff H. Room Acoustics, 6th ed. CRC Press; 2016.
6. Blauert J. Spatial Hearing: The Psychophysics of Human Sound Localization (Revised Edition). The MIT Press; 1996. doi: 10.7551/mitpress/6391.001.0001
7. Cecchi S, Carini A, Spors S. Room Response Equalization—A Review. Applied Sciences. 2017; 8(1): 16. doi: 10.3390/app8010016
8. Zölzer U. DAFX: Digital Audio Effects, 1st ed. Wiley; 2011. doi: 10.1002/9781119991298
9. Toole F. Sound Reproduction: The Acoustics and Psychoacoustics of Loudspeakers and Rooms, 1st ed. Focal Press; 2008.
10. Kuttruff H. Room Acoustics, 4th ed. Spon Press; 2000.
11. Sabine W. Collected Papers on Acoustics. Harvard University Press; 1922.
12. Anderson N. Acoustic Characterization of Rooms Using Directional and Omnidirectional Sources [Bachelor's Thesis]. Brigham Young University; 2022. pp. 2–3. Available online: https://brightspotcdn.byui.edu/d5/29/dba237634a4a80d523507798c5a1/nick-andersons22.pdf

13. Newell P. Recording Studio Design. Routledge; 2008. doi: 10.4324/9780080556055
14. Clarity Graph. Available online: https://www.roomeqwizard.com/help/help_en-GB/html/graph_clarity.html (accessed on 14 December 2025).
15. Sato H, Bradley JS. Evaluation of acoustical conditions for speech communication in working elementary school classrooms. The Journal of the Acoustical Society of America. 2008; 123(4): 2064–2077. doi: 10.1121/1.2839283
16. Mishan R, Fu X, Hingu C, et al. Analyzing frequency spectrum and Total Harmonic Distortion for high switching frequency operation of GaN-based filter-less multilevel cascaded H-bridge inverter. Advances in Electrical Engineering, Electronics and Energy. 2025; 11: 100906. doi: 10.1016/j.prime.2025.100906
17. Sundberg J. The Science of the Singing Voice. Northern Illinois University Press; 1987.
18. Lokki T, Pätynen J. Auditory Spatial Impression in Concert Halls. In: Blauert J, Braasch J (editors). The Technology of Binaural Understanding, Modern Acoustics and Signal Processing. Springer; 2020. pp. 173–202. doi: 10.1007/978-3-030-00386-9_7
19. Mi H. A Dynamic Reverberation Algorithm for Virtual Acoustic Rendering [Master's Thesis]. University of York; 2023. Available online: https://etheses.whiterose.ac.uk/id/eprint/36366/
20. Tao H, Li S, Wang X, et al. Analysis and Research on Spectrogram-Based Emotional Speech Signal Augmentation Algorithm. Entropy. 2025; 27(6): 640. doi: 10.3390/e27060640
21. Goetze S, Albertin E, Kallinger M, et al. Quality assessment for listening-room compensation algorithms. In: Proceedings of the 2010 IEEE International Conference on Acoustics, Speech and Signal Processing; 14–19 March 2010; Dallas, TX, USA. pp. 2450–2453. doi: 10.1109/ICASSP.2010.5496301
22. Bottalico P, Codino J, Cantor-Cutiva LC, et al. Reproducibility of Voice Parameters: The Effect of Room Acoustics and Microphones. Journal of Voice. 2020; 34(3): 320–334. doi: 10.1016/j.jvoice.2018.10.016
23. How the song "Rrjedh në këngë e đegërime" was born—Gjok Beci told the story. Available online: https://observerkult.com/si-lindi-kenga-rrjedh-ne-kenge-e-ligjerime-gjok-beci-rrefen-historine-vid/ (accessed on 14 December 2025). (in Albanian)
24. Stevens KN. Acoustic Phonetics. The MIT Press; 1999. doi: 10.7551/mitpress/1072.001.0001
25. Shaqiri K, Mehmeti B, Xhoni A, et al. The face that speaks, the voice that sings: the triad of orthodontics, music, and aesthetics. Rast Müzikoloji Dergisi. 2025; 13(3): 283–305. doi: 10.12975/rastmd.20251333
26. International Organization for Standardization (ISO). Acoustics — Measurement of Room Acoustic Parameters—Part 1: Performance Spaces (ISO 3382-1:2009). ISO; 2009. Available online: https://standards.iteh.ai/catalog/standards/sist/11128de8-0d0b-4a88-9ed2-707914dd2648/iso-3382-1-2009
27. Everest A, Pohlmann K. Master Handbook of Acoustics, 6th ed. McGraw-Hill; 2014.
28. Vorländer M. Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality, 1st edition. Springer; 2008. doi: 10.1007/978-3-540-48830-9
29. Cox T, D'Antonio P. Acoustic Absorbers and Diffusers Theory, Design and Application, 3rd ed. CRC Press; 2016.
30. Eyring CF. Reverberation Time In "Dead" Rooms. The Journal of the Acoustical Society of America. 1930; 1(2A): 217–241. doi: 10.1121/1.1915175
31. International Organization for Standardization (ISO). Acoustics—Measurement of Room Acoustic Parameters—Part 2: Reverberation Time in Ordinary Rooms. ISO; 2009.
32. Yousef AM, Hunter EJ. Sensitivity of Acoustic Voice Quality Measures in Simulated Reverberation Conditions. Bioengineering. 2024; 11(12): 1253. doi: 10.3390/bioengineering11121253
33. Franz S, Grewe T, Meyer BT, et al. Measurement and Analysis of Relevant Room Acoustic Parameters in Speech and Language Therapy Rooms and Evaluation of Their Influence on Objective Voice Quality Metrics. Journal of Voice. 2026. doi: 10.1016/j.jvoice.2025.12.012

Appendix A. Room acoustics: Additional measurements

High-resolution spectrogram in **Figure A1** shows pronounced harmonic smearing, modal ringing, and broadband energy dispersion in the untreated room.

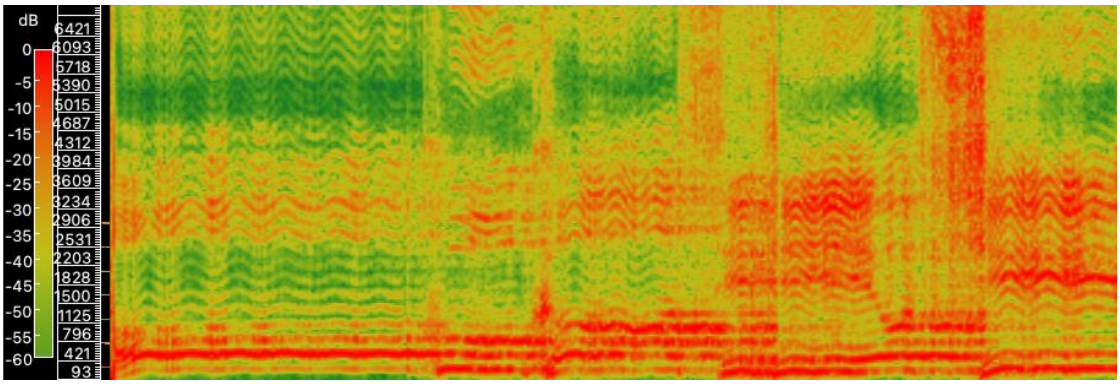

Figure A1. Extended Room Spectrogram—Untreated Room (Room B).

Detailed spectrogram in **Figure A2** confirms stable harmonic structure, minimal smearing, and uniform time–frequency behaviour.

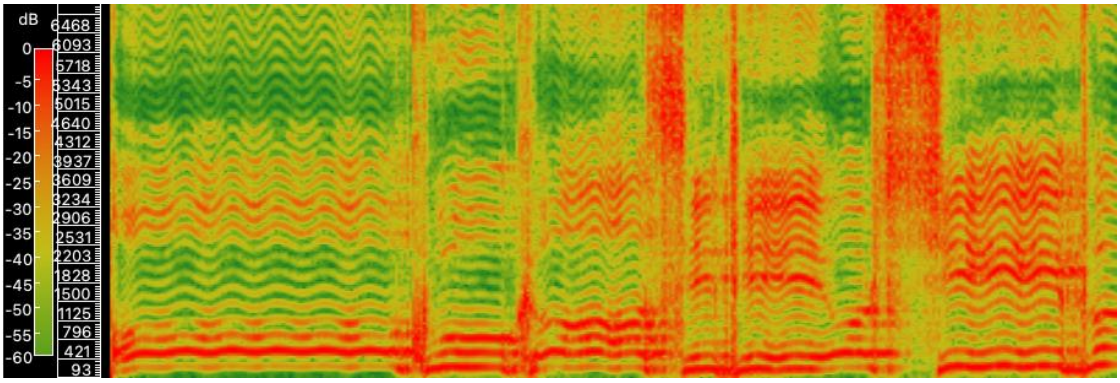

Figure A2. Extended Room Spectrogram—Treated Room (Room A).

Waveform analysis reveals prolonged decay, irregular temporal envelope, and persistent low-frequency buildup (**Figure A3**).

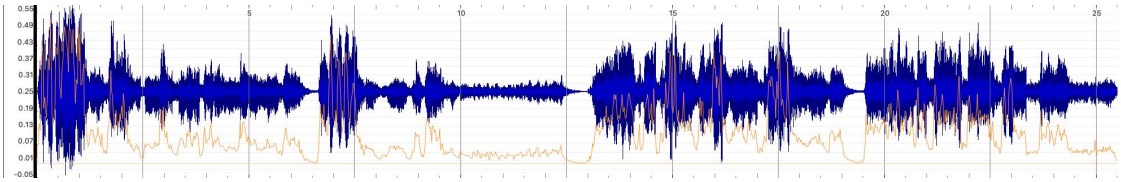

Figure A3. Waveform and Energy Envelope—Untreated Room (Room B).

Clean envelope with stable decay, indicates high acoustic control and minimal room colouration (**Figure A4**).

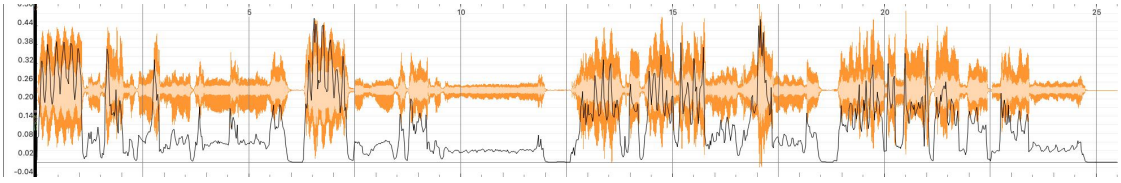

Figure A4. Waveform and Energy Envelope—Treated Room (Room A).

Figure A5 shows non-linear decay behaviour and extended low-frequency resonances typical of untreated small rooms.

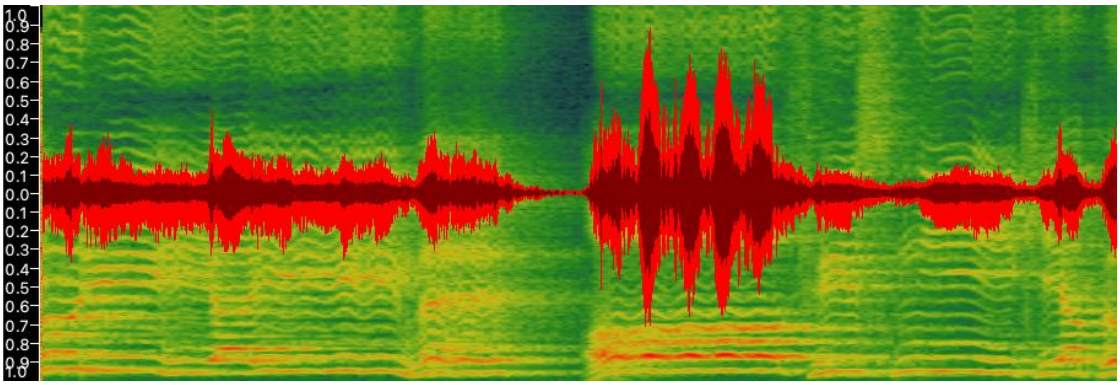

Figure A5. Energy Decay Curve (EDC)—Untreated Room (Room B).

Figure A6 illustrates smooth, rapid energy decay consistent with broadband absorption and controlled reflections.

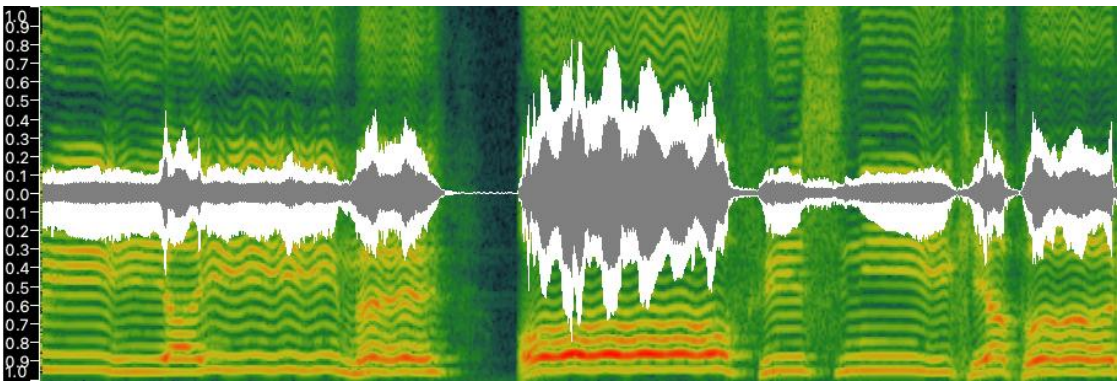

Figure A6. Energy Decay Curve (EDC)—Treated Room (Room A).

Second measurement confirms long decay tails and modal interference (**Figure A7**).

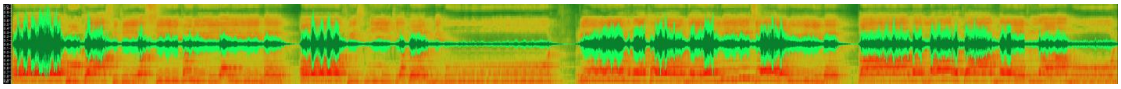

Figure A7. Additional EDC Visualisation—Untreated Room (Validation).

Supplementary decay graph demonstrates uniform decay slope and absence of modal ringing (**Figure A8**).

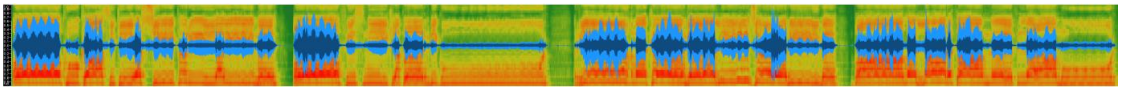

Figure A8. Additional EDC Visualisation—Treated Room (Validation).

Appendix B. Vocal signal analyses

High-detail vocal spectrogram in **Figure A9** shows clean harmonic definition and minimal time–frequency distortion.

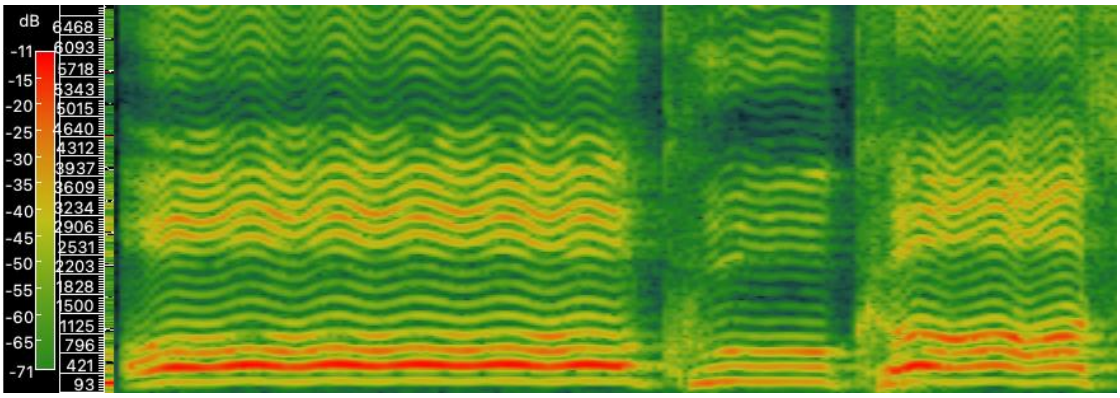

Figure A9. Vocal Spectrogram—Treated Room (Room A).

Displays fluctuations, formant instability, and increased spectral spread due to room coloration (**Figure A10**).

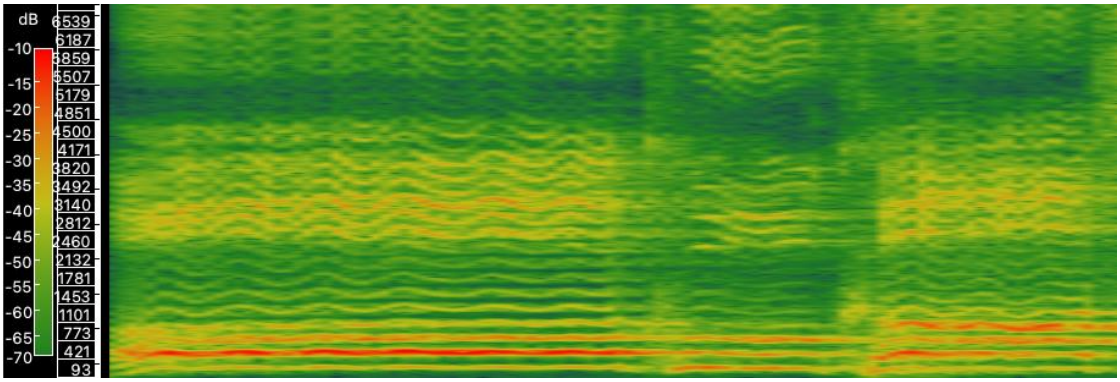

Figure A10. Vocal Vibration Spectrogram—Untreated Room (Room B).

Confirms early-reflection interference, modal masking, and unstable harmonic structure in the untreated room (**Figure A11**).

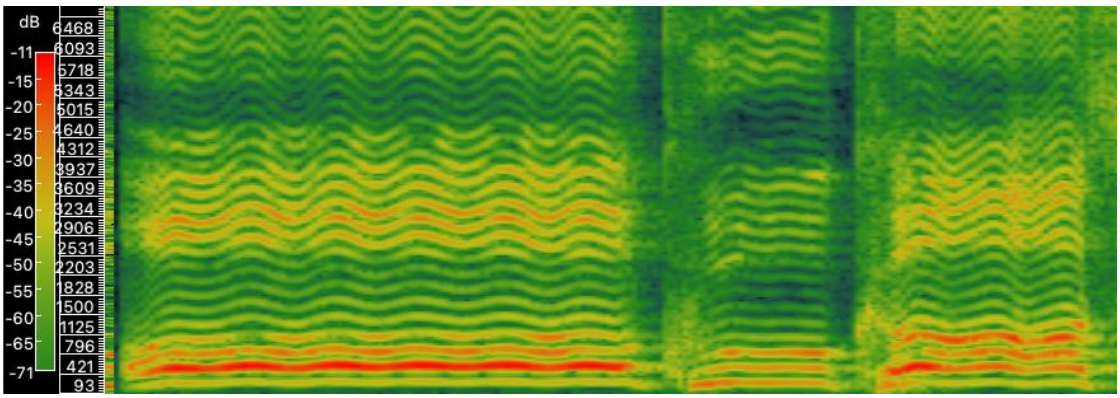

Figure A11. Additional Vocal Vibration Spectrogram—Untreated Room (Validation).