Research Square

Preprints are preliminary reports that have not undergone peer review. They should not be considered conclusive, used to inform clinical practice, or referenced by the media as validated information.

Trapped in a Soundproof Box: How CO2 Levels Influence Speech Quality Assessment and Response Behaviour

Leon Schreiber

leon.schreiber@tu-berlin.de

Technische Universität Berlin

Benedikt Schuh

Technische Universität Berlin

Kirill Shchegelskiy

Technische Universität Berlin

Wafaa Wardah

Technische Universität Berlin

Tuğçe Melike Koçak Büyüktaş

Technische Universität Berlin

Sebastian Möller

Technische Universität Berlin

Robert P. Spang

Technische Universität Berlin

Research Article

Keywords: Quality Assessment, CO, textsubscript{2} Exposure, Listening Tests, Environmental Effects, Perception

Posted Date: October 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11222803/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Trapped in a Soundproof Box: How CO_2 Levels Influence Speech Quality Assessment and Response Behaviour

Leon Schreiber[1*], Benedikt Schuh[1], Kirill Shchegelskiy[1], Wafaa Wardah[1,2], Tuğçe Melike Koçak Büyüktaş[1], Sebastian Möller[1,2], Robert P. Spang[1]

[1*]Quality and Usability Lab, Technische Universität Berlin, Berlin, Germany.
[2]German Research Center for Artificial Intelligence (DFKI), Berlin, Germany.

*Corresponding author(s). E-mail(s): leon.schreiber@tu-berlin.de;

Abstract

Standardised tests to assess speech quality, such as the listening tests of Recommendation P.800 of the International Telecommunication Union Telecommunication Standardization Sector (ITU-T), require controlled experimental environments to ensure reliable subjective ratings. However, poor ventilation in soundproof booths can lead to increased carbon dioxide (CO_2) concentrations, which may affect cognitive function and judgement. This study investigates whether CO_2 levels affect speech quality ratings and response behaviour in listening tests. We conducted a classic mean opinion score (MOS) listening experiment while continuously measuring CO_2 concentrations in a non-actively ventilated sound booth. Our analysis reveals three key findings: (1) Higher CO_2 levels significantly reduce response time, suggesting a shift towards more impulsive decision making. (2) Speech quality ratings are systematically affected, with poorer samples rated more favourably and higher quality samples rated less favourably at elevated CO_2 concentrations. (3) Response variability decreases under high CO_2 conditions, suggesting reduced sensitivity to quality differences. These findings highlight a previously overlooked factor in speech quality research, suggesting that ambient air quality may introduce bias into subjective ratings. We discuss the implications for experimental design and the need for controlled ventilation in perceptual studies to ensure data validity.

Keywords: Quality Assessment, CO_2 Exposure, Listening Tests, Environmental Effects, Perception

Main text: 3,451 words (excluding title page, references, figures and tables); 3 figures, 2 tables

1 Introduction

Controlled listening experiments remain a gold standard for evaluating human quality of experience in speech communication systems, particularly following the well-established P.800 methodology of the International Telecommunication Union Telecommunication Standardization Sector (ITU-T) [7]. At our lab, a newly acquired sound isolation cabin offered the opportunity to conduct such experiments in a dedicated, acoustically neutral environment. However, initial pilot tests revealed that the built-in ventilation system introduced a high level of continuous background noise, rendering it unsuitable for critical listening evaluations.

Faced with this constraint, we considered operating the cabin without mechanical ventilation for the duration of each session. During an early trial session, it quickly became apparent that the

air quality deteriorated noticeably after a short period of time. This raised an important, practical, and underexplored question: Does poor air quality—measurable via carbon dioxide (CO_2) as a proxy—influence participants' perceptual judgments in listening experiments?

1.1 Air Quality and Cognitive Performance

There is growing scientific evidence that elevated indoor CO_2 levels—even at concentrations previously considered safe—can impair cognitive performance in various tasks. Satish et al. [10] exposed participants to increasing concentrations of CO_2 (600, 1,000, and 2,500 parts per million, ppm) and found that decision-making performance declined significantly with increasing CO_2 levels, with statistically significant effects emerging in 6 out of 9 cognitive dimensions already at 1,000 ppm. Allen et al. [1] replicated and extended these findings in a simulated office environment with 24 participants, demonstrating further cognitive performance impairments at moderate (945 ppm) and high (1,400 ppm) CO_2 exposures.

While both studies focus on strategic multidimensional decision-making under extended exposure, the implications also raise concerns for any experimental protocols requiring concentration, perceptual discrimination, or evaluative decision-making—including speech quality listening tests, where participants are required to make repeated, fine-grained judgments under lab conditions.

1.2 CO_2 as an Indoor Air Quality Metric

CO_2 is widely recognized as a reliable proxy for human-generated air pollution and ventilation sufficiency. According to official guidelines [2] by the Austrian Federal Ministry for Climate Action, Environment, Energy, Mobility, Innovation and Technology, indoor CO_2 levels above 1,000 ppm are perceived as "unsatisfactory" by approximately 20% of people, and levels nearing 1,500 ppm are associated with thresholds of discomfort due to human metabolic emissions. Poorly ventilated spaces such as bedrooms and vehicles frequently exhibit CO_2 concentrations exceeding 1,400 ppm, with mean values above 2,000 ppm in taxis and buses [2]—concentrations which are comparable to those reached in our experimental setting.

1.3 Research Questions & Contributions

Motivated by these findings and observations, the present study investigated whether ambient CO_2 concentration affects subjective speech quality ratings in a controlled lab setting. Specifically, we aim to answer the following research questions:

- Does higher CO_2 concentration influence the perceived quality of speech samples?
- Are some stimuli (e.g., with especially good or poor quality) more affected than others?
- Does poor air quality attenuate rating precision or variability, e.g., through decision fatigue or perceptual blunting?
- What practical implications emerge for experimental design and data reliability in lab-based listening studies conducted without proper ventilation?

To address these questions, we equipped the cabin with a custom-built, high-resolution CO_2 monitoring system based on the MH-Z19 non-dispersive infrared (NDIR) sensor module and an ESP8266 microcontroller for real-time data streaming. We then conducted a classical ITU-T P.800-style listening test, staging 5,329 rating trials across 24 participants under varying CO_2 exposure. Our analysis combines classic perceptual quality modeling with signal processing techniques and linear mixed-effects modeling.

1.4 Extension over the Conference Paper

This article is an extended version of our QoMEX 2025 short paper [11]. Compared with the conference version, it adds an expanded review of related work on air quality and cognition and on CO_2 as an indoor air quality metric (Sects. 1.1 and 1.2), a complete description of the method (Sect. 2), the full statistical models for response times and ratings, including a translation of the rating model into predicted rating shifts (Sects. 3.1 and 3.2), new analyses of rating variability (Sect. 3.3), and an extended theoretical discussion and treatment

of limitations (Sect. 4). Figures 1 and 2, as well as Figs. S1 and S2 and Table S1 in Online Resource 1, are adapted or reproduced from the conference paper, as stated in their captions.

2 Methods

2.1 Participants

A total of $N = 24$ participants took part in the study, including 11 female, 12 male, 1 intersex/diverse person. The participants' ages ranged from 20 to 65 years, with a mean age of 36.29 years and a median age of 32 years. All participants were fluent in German (level C1 or above of the Common European Framework of Reference for Languages, CEFR), with self-reported normal hearing and no known cognitive or speech impairments. Each participant provided informed consent and received monetary compensation for their time. The study was approved by the Ethics Committee of Faculty IV (Ethik-Kommission der Fakultät IV), Technische Universität Berlin (reference number 322; application submitted on 24 February 2025).

2.2 Experimental Design and Procedure

The study followed a passive listening test protocol similar to the ITU-T P.800 standard [7]. Participants completed two consecutive experimental phases, separated by a 5 min break. The entire session took place in a sound-treated cabin with no active ventilation (mute labs Meetingbox QUATTRO).

First Phase: Participants rated 200 (8 sec to 12 sec each) isolated speech samples of varying quality sourced from a SwissQual AG dataset that was part of the training material for P.OLQA-based objective models, described in more detail in [15]. Each sample was presented once, and participants recorded their quality judgments using a mean opinion score (MOS) scale on an overflow range [8]. This phase lasted approximately 45 minutes.

Break Phase: Following the first part, the experimental software initiated a 5-minute break. During this time, participants could not interact with the system, but were informed that they could leave the cabin if desired. Most participants opened the door during this period, causing a natural ventilation of the cabin and a reduction in the CO_2 concentration.

Second Phase: After indicating their readiness to continue, participants proceeded to the second part, in which they rated 30 sec monaural speech segments from two-person conversations. Only one side of each conversation was audible, with segments selected to have at least 70 % active speech. These recordings were taken from a study environment operating over manipulated Web Real-Time Communication (WebRTC) connections [12], representing four distinct network degradation conditions (e.g., packet loss, reduced bandwidth, and high latency).

Participants performed the listening tests independently and were told they could pause and continue at their discretion. All instructions were given outside the cabin; a short test run was performed to ensure technical setup and comprehension before the session began.

2.3 CO_2 Measurement and Experimental Rationale

A key focus of the study was to examine the effect of elevated CO_2 concentration, which increases naturally during prolonged occupancy in closed rooms. Continuous CO_2 readings were captured using a custom-built sensor system based on an MH-Z19 module interfaced with an ESP8266 microcontroller (NodeMCU mini). Sensor readings were collected every two seconds and streamed via USB to a local laptop in the cabin. This allowed synchronous logging of CO_2 concentration alongside trial timestamps.

To experimentally separate time and CO_2 effects, the study was deliberately split into two phases with a break in between. This break increased the likelihood of participants opening the cabin door, effectively resetting or reducing the CO_2 concentration significantly. This setup introduced controlled variability in CO_2 dynamics across participants and trials, enabling subsequent modeling of its effect on perceived quality.

2.4 Measures

Two dependent variables were recorded: Speech Quality Rating (MOS): Participants used an overflow scale to rate perceived speech quality for

each trial. Response Time (s): The time interval between the end of audio playback and the initiation of the next trial, measured in seconds.

The primary independent variable was the CO_2 concentration (in ppm), time-aligned with each trial. The cumulative time elapsed since the first trial per participant, allowing for analysis of both cumulative exposure and CO_2 variation across time. All analyses were conducted in Python, using statsmodels for the linear mixed-effects models and SciPy for Levene's test; figures were created with matplotlib and seaborn.

2.5 Use of Generative AI

Generative AI (Claude, Anthropic) was used to assist in revising the manuscript for this journal version, including restructuring, drafting text passages from the authors' results and notes, and checking consistency and references. All analyses, results and interpretations are the authors' own; all AI-assisted text was reviewed and edited by the authors, who take full responsibility for the content.

3 Results

A total of 5,329 individual speech quality ratings were collected during the experiment. As expected, the level of CO_2 in the cabin increased progressively over time due to a lack of active ventilation. The rise in concentration continued until participants opened the door, allowing the cabin to ventilate naturally.

At the beginning of the experiment, the average CO_2 concentration was approximately 458 ppm. On average, it took 1190 sec (approximately 20 min) for the CO_2 level to reach 1,000 ppm, and around 2380 sec (approximately 40 min) to exceed 1,500 ppm. Fig. S1 in Online Resource 1 illustrates the increase in CO_2 concentration over time for each participant. Notably, there is a steep drop around the 2,900-second mark, corresponding to the moment when most participants opened the cabin door to take a break.

3.1 Effect of CO_2 on Response Time

To investigate the influence of ambient CO_2 concentration on the speed at which human raters assess speech quality, a linear mixed-effects model was fitted using participant ID as a random intercept to account for inter-rater variability. The fixed effects included the CO_2 concentration (continuous, in ppm) and relative time since experiment start (continuous, in sec). The model was estimated using restricted maximum likelihood (REML).

CO_2 concentration had a statistically significant negative effect on response time ($\beta = -0.005$, SE $= 0.001$, $z = -4.748$, $p < .001$), indicating that higher CO_2 levels were associated with faster rating responses. Relative experiment time also positively predicted response time ($\beta = 0.005$, SE $= 0.000$, $z = 15.866$, $p < .001$), suggesting a gradual slowing of responses over time. The intercept was 4.095 (SE $= 1.136$, $z = 3.606$, $p < .001$), and the variance of the random intercept was 0.000 (SD $= 0.057$).

To further illustrate the relationship between CO_2 and rating speed, we examined the average response times across five ordinal CO_2 bucket levels (Low, Mid-Low, Mid, Mid-High, High) with each bucket containing an equal number of ratings. The buckets covered the following CO_2 ranges (mean $\pm$ SD in parentheses): Low 406–706 ppm (581.15±84.004), Mid-Low 707–948 ppm (831.97± 70.065), Mid 949–1187 ppm (1062.86 ± 67.661), Mid-High 1188–1397 ppm (1292.54 ± 60.616) and High 1398–2202 ppm (1627.50 ± 197.146). As seen in Table S1 in Online Resource 1, response time decreased with higher CO_2 levels, supporting the negative association found in the mixed model.

3.2 Effect of CO_2 on Subjective Speech Quality Ratings

To investigate the influence of ambient air quality on speech perception, we analyzed the relationship between participants' MOS and the CO_2 concentration in the sound cabin.

Figure 1 shows the distribution of MOS values across CO_2 concentration levels. Under conditions of good air quality (low CO_2), participants made wide use of the full rating scale. As CO_2 accumulated in the cabin, this distribution narrowed: ratings clustered more closely around the midpoint of the scale. This gravitation toward central ratings indicates a dampening of participants' ability—or willingness—to make nuanced and fine-grained quality distinctions under degraded air quality conditions.

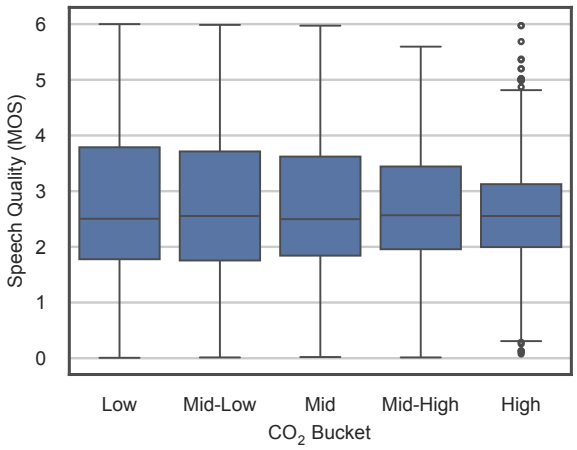

Fig. 1 Distribution of MOS ratings of all samples in the test across different CO_2 concentration levels (bucketed). Adapted from [11], © 2025 IEEE

To formally analyze the effect, we again fitted a linear mixed-effects model predicting subjective ratings as a function of:

- the continuously logged CO_2 concentration,
- the MOS of each stimulus across all participants,
- their two-way interaction,
- and time spent in the cabin.

A random intercept was fitted for each participant to account for inter-individual rating biases. The results are summarized in Table 1.

Table 1 Effects of CO_2, MOS, and time on subjective ratings

Predictor	Coef.	SE	z	p
Intercept	-0.515	0.150	-3.43	$< .001$
CO_2	4.92×10^{-4}	9.35×10^{-5}	$+5.26$	$< .001$
Avg. Quality	$+1.190$	0.037	$+32.50$	$< .001$
$CO_2\times$Qual.	-1.93×10^{-4}	3.23×10^{-5}	-5.99	$< .001$
Relative Time	2.15×10^{-5}	9.61×10^{-6}	$+2.24$	.025

All fixed effects in the model were statistically significant. Because CO_2 is measured in raw ppm and spans nearly 1,800 units in our data, the CO_2 and CO_2 × Avg. Quality coefficients in Table 1 are necessarily small in magnitude. The positive main effect of CO_2 indicates a subtle increase in MOS ratings as air quality worsens. However, this main effect is qualified by a strong and significant interaction between CO_2 concentration and the average perceived quality of each sample. This finding suggests a compression of ratings: mediocre-quality samples remained relatively stable across different CO_2 levels, whereas extreme-quality samples were rated differently depending on air quality. Table 2 translates this interaction into rating units, showing the model-predicted MOS shift across the full observed CO_2 range for the five quality tiers used in Fig. 2.

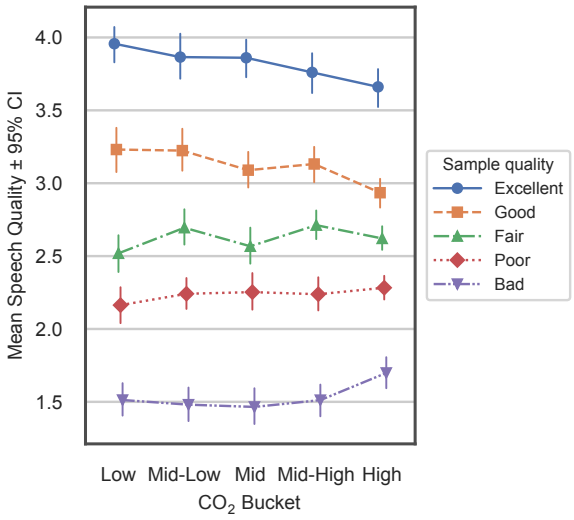

Fig. 2 Interaction effect between CO_2 concentration and sample quality on participant MOS ratings. Samples have been grouped into five equally sized categories based on their MOS. High-quality samples receive lower MOS ratings under high CO_2; poor-quality samples receive inflated ratings as ambient CO_2 increases. Adapted from [11], © 2025 IEEE

Table 2 Model-predicted change in MOS across the observed CO_2 range (~400–2,200 ppm), by sample quality tier

Quality Tier	Avg. MOS	Predicted ΔMOS
Bad	1.55	$+0.35$
Poor	2.23	$+0.11$
Fair	2.61	-0.02
Good	3.13	-0.20
Excellent	3.86	-0.46

Figure 2 visualizes this interaction. Under low CO_2 conditions, ratings closely reflect the inherent quality of the stimulus: high-quality samples are

rated highly, while poor-quality samples receive low ratings. As air quality degrades (i.e., CO_2 increases), this relationship weakens. High-quality samples are rated significantly worse, while poor-quality samples receive more favorable ratings. This compression of scaling indicates a diminished sensitivity or discrimination ability—participants' ability to distinguish between excellent and poor audio appears to degrade under higher CO_2 concentrations.

3.3 Effect of CO_2 on Rating Variability

In addition to analyzing central tendencies of subjective ratings, we examined whether air quality influenced the consistency of listeners' judgments. Specifically, we asked whether increased CO_2 concentrations led to reduced variability in MOS.

Fig. S2 in Online Resource 1 summarizes the standard deviation of speech quality ratings across CO_2 concentration levels (bucketed). The variability of ratings decreases markedly as air quality deteriorates, with the highest standard deviation observed under the lowest CO_2 condition (*Low*: $\sigma = 1.26$), and the lowest under the highest CO_2 condition (*High*: $\sigma = 0.96$). The monotonic trend suggests a consistent reduction in inter-rater variability as ambient air quality declines.

This observation was statistically confirmed using Levene's test for equality of variances across the CO_2 conditions, which returned a highly significant result ($F = 32.06$, $p < .001$). This indicates that the variances differ significantly between CO_2 bins, with a substantial compression of spread under poorer air quality conditions.

To further investigate this phenomenon at the individual participant level, we modeled within-subject variability by computing the log-transformed residuals per trial, capturing the degree to which individual ratings deviated from each participant's average. A linear mixed-effects model with random intercepts for participants showed a significant negative effect of CO_2 concentration on residual variability ($\beta = -0.000$, $p < .001$), confirming that participants' responses became less variable as CO_2 levels increased.

To complement this analysis, we also evaluated a standard consistency metric relating the standard deviation of opinion scores (SOS) to the MOS [6]. Figure 3 plots the standard deviation of opinion scores (SOS) against the MOS for each of the 50 test conditions of the dataset, grouped by CO_2 bucket. Solid lines show SOS hypothesis fitting functions, where the SOS parameter a is obtained by minimizing the least squared errors. Lower SOS parameter values correspond to lower diversity in raters' judgments. As shown in the plot, responses under higher CO_2 conditions demonstrate both a reduction in rating spread and a lower fitted SOS slope, consistent with a monotonic narrowing of user responses. Overall, SOS parameters were consistent with values typically reported in the speech quality literature (0.17–0.25) [5], while more extreme CO_2 levels resulted in significantly more stable and less variable ratings outside of the common range.

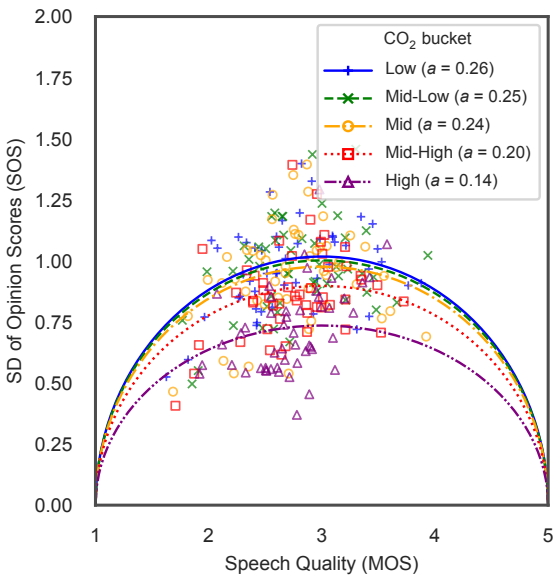

Fig. 3 SOS-MOS plot across different CO_2 conditions. Each marker represents a test condition (MOS, SD). Lines indicate fitting functions according to the SOS hypothesis with the SOS parameter a specified in the legend. Rating diversity decreases with increasing CO_2

4 Discussion

The results of this study indicate that increasing ambient CO_2 concentration affects listener behavior in perceptual speech quality assessment. In particular, we observed a compression effect: ratings under higher CO_2 exposures clustered closer to the center of the scale, with high-quality samples being rated less favorably and low-quality

ones more leniently. This distortion of perceptual judgments challenges the assumption that subjective ratings under such controlled conditions reflect only content-related perceptual differences, and raises questions about how ambient factors influence data quality more broadly.

This compression is theoretically supported by several strands of psychological research. From a cognitive load theory perspective [9, 13], higher environmental load—such as reduced air quality—can impose additional extraneous cognitive burden, narrowing the resources available for task-specific processing. Applied to the task of fine-grained speech quality judgment, participants may default to mid-scale responses as a heuristically simpler strategy. This simplification effect is compatible with findings in educational psychology, where learners under cognitive strain abandon effort-intensive strategies in favor of broader generalizations or shortcuts.

A complementary explanation stems from self-regulation theory. Under sustained environmental discomfort, such as humid or stale air, the active self becomes depleted, as described in the ego depletion literature [3, 14]. This depletion reduces individuals' capacity to exert deliberate effort over time, particularly in repeated decision-making or evaluative tasks. Rather than maintaining precision when rating each stimulus, listeners may adopt a more emotionally neutral stance, giving ratings that reflect neither strong satisfaction nor dissatisfaction.

From a perceptual standpoint, the results are also consistent with predictions from signal detection theory [4], whereby increased internal noise—whether cognitive or sensory—reduces discriminability between stimuli. If CO_2-induced fatigue raises such internal uncertainty, listeners may find it harder to reliably tell apart nuanced differences in audio quality. The resulting over-reliance on the middle of the scale closely mimics the reduced sensitivity outcomes signal detection models predict under high-noise conditions.

The observed decrease in rating variability adds further weight to this interpretation. Listeners appear less decisive or expressive in their judgments, suggesting not only a perceptual effect, but also a metacognitive one: under degraded environmental conditions, participants may become less confident in their judgments, or less motivated to engage with the task. Intriguingly, response time analyses showed that decision speed increased with CO_2, suggesting even further alignment with cognitive shortcut-taking and response simplification. Whether this indicates impulsivity or strategic resource-saving remains open for further inquiry.

Together, these theoretical lenses converge on a single behavioral pattern: when environmental stress or discomfort taxes cognitive resources, humans adapt by perceiving less distinctly, judging less precisely, and deciding more simplistically.

4.1 Limitations and Future Work

While our findings are compelling, several limitations should be acknowledged.

First, unlike previous studies [1, 10], CO_2 levels in this study were not experimentally manipulated or held constant across conditions. Instead, they evolved naturally over time due to participant respiration in a closed environment. While this increases ecological validity, it limits causal precision and introduces potential confounds between CO_2 and time-on-task effects—which we attempted to control statistically and through our experimental design.

Second, the listening task involved passive quality rating, which may impose lower cognitive demands than the complex decision-making tasks used in prior CO_2 studies. Future work should investigate whether similar perceptual and behavioral distortions emerge in other experimental paradigms, such as conversational interaction or other collaborative tasks.

Third, although our CO_2 sensor provided high temporal resolution, it does not directly capture other indicators of indoor air quality (e.g., temperature, humidity, volatile organic compounds). It remains possible that other environmental factors contributed to the observed effects.

Despite these limitations, our findings raise important methodological considerations. Future work should replicate these results across diverse participant pools, task paradigms, and controlled CO_2 exposure levels. Given the potential for uncontrolled environmental variables to influence perceptual judgments, we advocate for the inclusion of real-time ambient monitoring—including CO_2—as a standard data stream in lab-based experiments.

5 Conclusion

This study set out to examine whether CO_2 levels—an often-overlooked environmental variable—can influence subjective judgments in perceptual listening tests. Across more than five thousand quality ratings collected in an enclosed, non-actively ventilated sound cabin, we found compelling evidence that the concentration of CO_2 in the room significantly affected both the outcome and character of listener responses.

As CO_2 levels rose during the course of each experimental session—frequently surpassing 1,500 ppm within 45 minutes—we observed systematic distortions in listeners' ratings. Under higher concentrations, participants tended to rate good quality samples less favorably, while poor quality samples were evaluated more leniently. This pattern suggests a narrowing of perceptual sensitivity, producing a rating compression. Consistent with this finding, the variability of responses across raters also decreased, indicating that under degraded air quality, listeners provided less differentiated, less diverse evaluations.

These results have important methodological implications for speech and audio research: ambient environmental factors such as air quality are not neutral backdrops to perceptual experiments, but active contributors to cognitive and perceptual performance. In long-form listening tests conducted in close-to-sealed environments, it is essential to control air quality, either through ventilation, time-limiting protocols, or environmental sensing, if we are to ensure the reliability and interpretability of subjective data.

Ultimately, our findings point to a simple but often overlooked truth: ensuring high-quality experimental data in perceptual research may require not only careful stimulus design and rigorous procedures, but also clean, breathable air.

Supplementary information. Online Resource 1 (ESM_1.pdf) contains Fig. S1 (CO_2 concentration over time for each participant), Table S1 (response time per CO_2 bucket) and Fig. S2 (standard deviation of the ratings per CO_2 bucket).

Statements and Declarations

Funding. No funding was received for conducting this study.

Competing interests. The authors have no relevant financial or non-financial interests to disclose.

Ethics approval. The study was approved by the Ethics Committee of Faculty IV (Ethik-Kommission der Fakultät IV), Technische Universität Berlin (reference number 322; application submitted on 24 February 2025).

Consent to participate. Informed consent was obtained from all individual participants included in the study.

Data availability. The dataset and analysis scripts supporting the findings of this study are available from the corresponding author on reasonable request. They will be made publicly available under a Creative Commons Attribution 4.0 International (CC BY 4.0) license upon publication.

Author contributions. R. P. Spang conceived the study. L. Schreiber collected the data. K. Shchegelskiy performed the SOS–MOS analysis. All authors contributed to the preparation of the manuscript, and all authors read and approved the final manuscript.

References

[1] Allen JG, MacNaughton P, Satish U, et al (2016) Associations of cognitive function scores with carbon dioxide, ventilation, and volatile organic compound exposures in office workers: A controlled exposure study of green and conventional office environments. Environmental Health Perspectives 124(6):805–812. https://doi.org/10.1289/ehp.1510037

[2] Arbeitskreis Innenraumluft im BMK (2024) Richtlinie zur Bewertung der Innenraumluft – Kohlenstoffdioxid als Lüftungsparameter. Bundesministerium für Klimaschutz, Umwelt, Energie, Mobilität, Innovation und Technologie (BMK), Wien, URL https://www.bmluk.gv.at/dam/jcr:e791f304-7dcf-4783-bc5a-af25daa8afb3/

Richtwert_CO2_UA.pdf, accessed 29 September 2026

[3] Baumeister RF, Bratslavsky E, Muraven M, et al (2018) Ego depletion: Is the active self a limited resource? In: Self-Regulation and Self-Control. Routledge, p 16–44, https://doi.org/10.4324/9781315175775-1

[4] Hautus MJ, Macmillan NA, Creelman CD (2021) Detection Theory: A User's Guide. Routledge, New York, https://doi.org/10.4324/9781003203636

[5] Hoßfeld T, Schatz R, Egger S (2011) SOS: The MOS is not enough! In: 2011 Third International Workshop on Quality of Multimedia Experience (QoMEX). IEEE, pp 131–136, https://doi.org/10.1109/QoMEX.2011.6065690

[6] Hoßfeld T, Heegaard PE, Varela M, et al (2016) QoE beyond the MOS: an in-depth look at QoE via better metrics and their relation to MOS. Quality and User Experience 1(1):2. https://doi.org/10.1007/s41233-016-0002-1

[7] ITU-T Recommendation P.800 (1996) Methods for subjective determination of transmission quality. International Telecommunication Union, Geneva

[8] Köster F, Guse D, Wältermann M, et al (2015) Comparison between the discrete ACR scale and an extended continuous scale for the quality assessment of transmitted speech. In: Fortschritte der Akustik – DAGA 2015, pp 150–153

[9] Paas F, Renkl A, Sweller J (2003) Cognitive load theory and instructional design: Recent developments. Educational Psychologist 38(1):1–4. https://doi.org/10.1207/S15326985EP3801_1

[10] Satish U, Mendell MJ, Shekhar K, et al (2012) Is CO_2 an indoor pollutant? Direct effects of low-to-moderate CO_2 concentrations on human decision-making performance. Environmental Health Perspectives 120(12):1671–1677. https://doi.org/10.1289/ehp.1104789

[11] Schreiber L, Schuh B, Shchegelskiy K, et al (2025) Cognitive fog under elevated CO_2: The influence of air quality on speech quality assessment and response behaviour. In: 2025 17th International Conference on Quality of Multimedia Experience (QoMEX). IEEE, Madrid, Spain, pp 1–4, https://doi.org/10.1109/QoMEX65720.2025.11219970

[12] Spang RP (2024) Operationalization of the IAE model entities: Relevant dimensions in the context of this work. In: Individualized Quality of Experience Estimation in Audiovisual Communication. T-Labs Series in Telecommunication Services, Springer, Cham, p 43–64, https://doi.org/10.1007/978-3-031-67551-5_4

[13] Sweller J (1988) Cognitive load during problem solving: Effects on learning. Cognitive Science 12(2):257–285. https://doi.org/10.1207/s15516709cog1202_4

[14] Vohs KD, Baumeister RF, Schmeichel BJ, et al (2018) Making choices impairs subsequent self-control: A limited-resource account of decision making, self-regulation, and active initiative. In: Self-Regulation and Self-Control. Routledge, p 45–77, https://doi.org/10.4324/9781315175775-2

[15] Wardah W, Vergari M, Spang RP, et al (2025) Perceptual drift in speech quality: Re-assessing listener judgments a decade later. In: DAS|DAGA 2025 – 51st Annual Meeting on Acoustics, Copenhagen, pp 1287–1290, URL https://pub.dega-akustik.de/DAS-DAGA_2025/files/upload/paper/457.pdf

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ESM1.pdf