forum acusticum 2026
8-12 September · Graz, Austria

# ANNOYANCE RATINGS OF HEAT PUMP NOISE IN DIFFERENT AURALISED SCENARIOS

Lara Stürenburg, Lukas Aspöck, Janina Fels

*Institute for Hearing Technology and Acoustics, RWTH Aachen University, Germany*

This paper is a revised version of the authors' submitted manuscript that was peer-reviewed by at least two anonymous reviewers.



## Abstract

Air-to-water heat pumps are increasingly used for heating, making their acoustic emissions an important factor in public acceptance. The perceived annoyance of heat pump noise depends not only on the sound characteristics of the unit but also on the installation location. The aim of this study was to investigate how simulated propagation scenarios (*garden*, *living room* with tilted window, and *free-field* reference), loudness conditions (*original loudness* and *equal loudness*), and contextual framing (*own heat pump* vs *neighbour's heat pump*) affect the short-term annoyance of heat pump noise. A listening experiment was conducted in which participants rated the annoyance of heat pump stimuli using an 11-point rating scale. The stimuli are from recordings of one heat pump in two different operating conditions. The results show clear differences in the annoyance ratings between the auralised scenarios, with lower ratings observed in the *living room* scenario than in the *garden* scenario. Equalising the loudness of the stimuli and the contextual framing had only minor or no effects on the ratings, and no clear influence of the operating condition was observed. Overall, the results indicate that the auralised scenario has the strongest influence on the perceived short-term annoyance of heat pump noise.

**Keywords:** *annoyance, heat pump, auralisation, listening experiment*

## 1. Introduction

In recent years, heat pumps have played an increasingly important role in supplying heat to buildings. The acoustic emissions of air-to-water heat pumps have a decisive influence on their acceptance. Not only do the fan(s) and compressor contribute to noise generation and perception, but also the installation configuration of the components, the installation location, and the operating condition. Acun et al. [1] conducted a listening experiment to assess human responses to air-source heat pump noise under different operating conditions, source distances, and background noise levels. They found no significant difference in the ratings between operating states. This deviates from the findings of Stürenburg et al. [2], which found differences in annoyance ratings for different operating conditions. In their experiment [2], however, only two stimuli per operating condition were examined, which were also based on the same recordings [3] as the stimuli used in this study, but with different sound propagation simulations. Most studies regarding heat pump noise either recorded sound stimuli in climate chambers [4] or did field measurements of operating heat pumps in residential areas [5].

In the study presented here, the recordings are of an air-to-water heat pump operated in a hemi-anechoic chamber and are auralised afterwards for a listening experiment in different simulated propagation scenarios. This approach yields stimuli that include realistic sound propagation effects while preserving full experimental control. In previous work [6], we conducted a more in-depth analysis of the stimulus characteristics for a limited selection of stimuli which are also evaluated in the study of this paper. It was found that the perceived loudness and preference can not be assessed simply through sound pressure level measurements when analysing heat pump noise. To get a broader insight into the annoyance of the heat pump noise and the impact of the receiver position and on the contextual framing of the listener, the study presented here investigates the effect of different listener positions by auralising an outdoor garden scenario and an indoor living room scenario, and the effect of different instructions towards the participants regarding the origin of the stimuli (*own heat pump* vs *neighbour's heat pump*). A reference scenario represents the free-field measurements of the heat pump recordings. Further, the listening experiment was conducted under two different loudness conditions to gain insight into the

***Corresponding author:*** *lara.stuerenburg@akustik.rwth-aachen.de.*

influencing factor of loudness. The participants rated the annoyance of the 3.5 s-long heat pump stimuli using an 11-point rating scale. Due to the short stimulus presentations in the listening experiment, the annoyance assessed in this study refers to short-term annoyance, and should not be considered equivalent to the long-term annoyance assessed in observational social survey settings.

## 2. Methods

### 2.1. Stimuli

Twenty-six heat pump recordings were chosen as the stimuli for the listening experiment, further details can be found in the technical report [3]. The recordings used in the present study are the recordings 1 to 26 of [3] – identical numbering is used in this paper. Additionally, a reference sound was included in the listening experiment. This reference stimulus was generated by filtering white noise to approximate the specific loudness of the measurement recordings. All signals were cut to a duration of 3 s. No background noise was added to the listening experiment.

### 2.2. Loudness Conditions

To get an impression of how much loudness influences the annoyance ratings of the stimuli, a set of stimuli was prepared in which all stimuli were set to an equal loudness. The chosen loudness value was the mean loudness of all 26 heat pump recordings. The stimuli were set to equal loudness before the auralisation, see Section 2.3. The listening experiment was conducted for all stimuli in both conditions: *original loudness* and *equal loudness*.

### 2.3. Simulations

Since outdoor units of heat pumps are typically not operating in free-field conditions, the sound propagation of a typical installation situation was simulated for two listening positions. All simulations were performed using the ray-based simulation software RAVEN [7], [8]. The propagation of the recorded heat pump noise was binaurally auralised, accounting for the directivity of the heat pump and the interaction of the sound with the surrounding environment. The directivity of the heat pump was computed from the measured directivity data of the respective heat pump [3]. The signals were set to a duration of 3.5 s to account for the reverberation generated by the acoustic simulation. Two acoustic scenarios were simulated: *garden* and *living room*. The heat pump is at the same position in both scenarios, but the position of the receiver differs. The third scenario is *free-field*, a reference situation including only the direct sound of the heat pump.

#### Scenario 1: Garden

In this scenario, the heat pump is located near the wall of the neighbouring building, and the listening position is approximately 13.1 m away from the heat pump with an additional lateral displacement of 0.4 m on a terrace at the rear of a typical detached house, see Fig. 1.

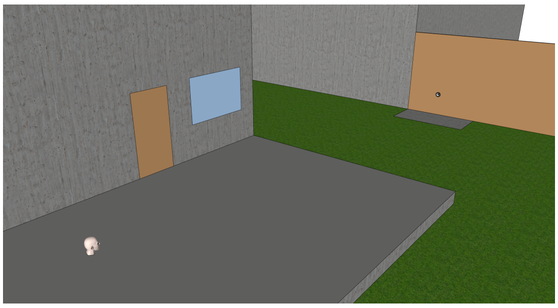

**Figure 1:** Simulation of the outdoor environment. The heat pump is located on the wall of the neighbouring building. For the *garden* scenario, the listening position is 13.1 m away from the heat pump with an additional lateral displacement of 0.4 m, depicted by the head. For the *living room* scenario, the listening position is behind the window shown in the sketch.

#### Scenario 2: Living room

In this scenario, the heat pump is at the same position as in the *garden* scenario, but the listening position is behind the window shown in Fig. 1, in a living room, see Fig. 2. The window is tilted, and the listening position is 2.0 m away from the window, slightly facing the window. The binaural impulse response was simulated in three steps: 1) impulse response from the heat pump to the window; 2) consideration of a sound-insulation measure for the tilted window; 3) binaural room impulse response from the window to the receiver position in the living room. The tilted window was modelled using a simple approximation with a low-pass filter (first-order Butterworth) with a cutoff frequency of 1000 Hz and an additional frequency-independent attenuation of 8 dB.

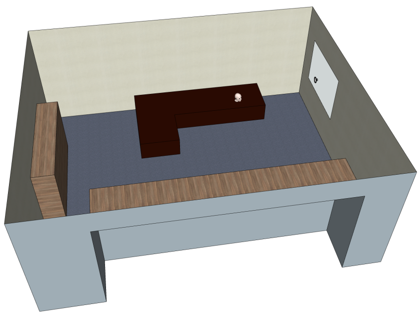

**Figure 2:** Simulation of the living room for the *living room* scenario. The listening position (head) is 2 m behind the tilted window.

#### Scenario 3: Free-field

This scenario serves as a reference scenario. No environment was simulated here; instead, only the direct sound for a receiver positioned 3 m in front of the heat pump was simulated. This scenario corresponds to the situation in the hemi-anechoic chamber during the recordings of the heat pump.

| Loudness | Scenario | $N_{\mathrm{min}}$ [sone] | $N_{\mathrm{max}}$ [sone] | $N_{\mathrm{mean}}$ [sone] |
|---|---|---|---|---|
| Original | *Garden* | 7.8 | 15.2 | 10.1 |
| | *Living room* | 0.9 | 2.9 | 1.5 |
| | *Free-field* | 12.9 | 24.0 | 16.0 |
| Equal | *Garden* | 9.4 | 11.7 | 10.1 |
| | *Living room* | 1.3 | 1.6 | 1.4 |
| | *Free-field* | 15.6 | 17.6 | 16.1 |

**Table 1:** Loudness values across all stimuli given in sone for the left ear: the minimum loudness $N_{\mathrm{min}}$, the maximum loudness $N_{\mathrm{max}}$, and the mean loudness $N_{\mathrm{mean}}$.

### 2.4. Listening Experiment

The listening experiment consisted of twelve blocks, with blocks seven to twelve being repetitions of blocks one through six. The six blocks were the scenarios *garden*, *living room*, and *free-field* in both loudness conditions *original loudness* and *equal loudness*. Even though the loudness of all source signals was set equal for the *equal loudness* condition before auralisation, the resulting auralised stimuli showed small variations in loudness. The corresponding loudness values for each loudness condition and scenario are given for the left ear in Table 1. The reported values $N_{\mathrm{min}}$ and $N_{\mathrm{max}}$ correspond to the individual stimuli with the lowest and highest loudness, respectively, and $N_{\mathrm{mean}}$ represents the mean loudness across all stimuli within each scenario. It can be seen that the ranges in loudness between stimuli are smaller in the *equal loudness* condition, whereas $N_{\mathrm{mean}}$ is identical within $\pm 0.1$ sone in both loudness conditions.

The participants were asked "How annoying is this sound?", and had to rate the sounds on a scale from 0 (not at all annoying) to 10 (extremely annoying). The order of the blocks was balanced across participants using random Latin squares, but it was always the six blocks first, then their repetition. The order of the repetitions was different from the order of the first six blocks. Thirty-one participants took part in the listening experiment (7 women, 24 men), aged between 21 and 39 years (mean age: 26.6 years). All participants had normal hearing ($\leq$ 20 dB(HL)), as confirmed by audiometric screening. Participation was voluntary, and written informed consent was obtained from all participants. Due to technical issues resulting in incomplete data, five participants had to be excluded from the analyses. Furthermore, to ensure data quality, participants were excluded from the analyses after assessing test-retest reliability using Spearman's rank correlation coefficient $\rho$. The coefficients were calculated by considering all annoyance ratings for each participant across all blocks. Participants with weak correlation $\rho < 0.2$ were considered inconsistent in their ratings, and all their data were excluded from the subsequent analyses. Five participants were found to be inconsistent, yielding a final dataset of 21 participants. During the listening experiment, participants sat in a listening booth, and the heat pump stimuli were presented via Sennheiser HD 650 headphones.

### 2.5. Ownership of the Heat Pump

The participants were divided into two groups: *own heat pump* and *neighbour's heat pump*. They were instructed that the sounds originate either from their *own* heat pump or from their *neighbour's* heat pump. During the experiment, before the start of each block, they were reminded of this fact. In the final dataset of 21 participants, 9 participants were in the *own heat pump* group and 12 participants were in the *neighbour's heat pump* group. The sample size for each condition effectively doubles since each individual test and retest rating is regarded as an independent observation.

## 3. Results

### 3.1. Test-Retest

Fig. 3 shows six heat maps for the test-retest of the annoyance ratings for each block. On the left side (blue) are the blocks using the *original loudness* stimuli, and on the right (red) are the blocks using *equal loudness* stimuli. From top to bottom are the scenarios *garden*, *living room*, and *free-field*. As a visual aid, a 1:1 agreement line in black is plotted on top of the heat map. With a total of 27 stimuli and 21 participants, there are 567 test-retest pairs. The colour map is fixed by 56, as a maximum agreement of 56 times was reached at the *free-field* scenario for the annoyance rating of 8.

### 3.2. Differences in scenarios and loudness

In Fig. 3, it can be seen how the participants rated the stimuli across the different scenarios. To get a better overview and comparison of them, the median and interquartiles are shown in Fig. 4 for the three scenarios (top to bottom: *garden*, *living room*, and *free-field*) and for the two loudness conditions (blue: *original loudness*, red: *equal loudness*), across all participants and stimuli. It can be seen that the ratings across scenarios vary in their median ratings; the *garden* scenario has a median annoyance rating of 6, the *living room* scenario has a lower median annoyance rating of 4, and the *free-field* scenario has the highest median annoyance rating of 7. The interquartile ranges (IQRs) are between 2 and 3 rating points. There is no difference in the median ratings between stimuli in the *original loudness* condition and in the *equal loudness* condition. Only the IQRs vary in the *garden* and *free-field* scenarios. Pairwise comparisons between the scenarios were performed with Wilcoxon signed-rank tests. Each test and retest annoyance rating across participants and stimuli was considered an independent observation, effectively doubling the sample size for each condition. Effect sizes were expressed as $r$. The results of the statistical comparisons are summarised in Table 2. Significant differences ($p < 0.001$) were found between all scenarios within both loudness conditions; effect sizes were huge ($r > 0.7$) for the comparisons involving the *living room* scenario, and large ($r > 0.3$) between the *garden* and *free-field* scenarios. In

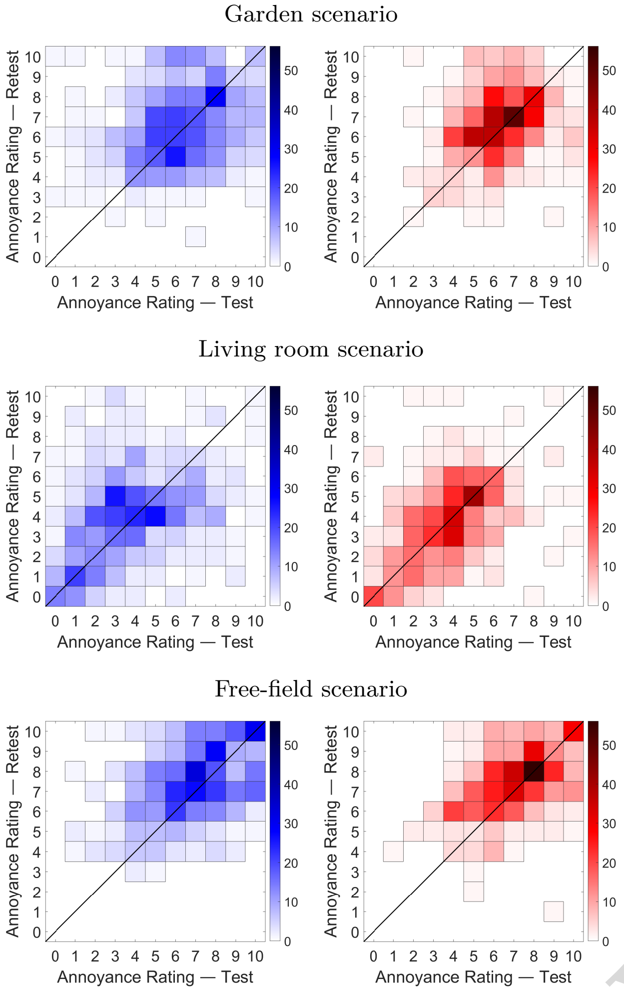


**Figure 3:** Heat map of the test-retest annoyance ratings. Top to bottom: *garden* scenario, *living room* scenario, *free-field* scenario. Left: *original loudness*, right: *equal loudness*. The colour scale represents the number of occurrences.

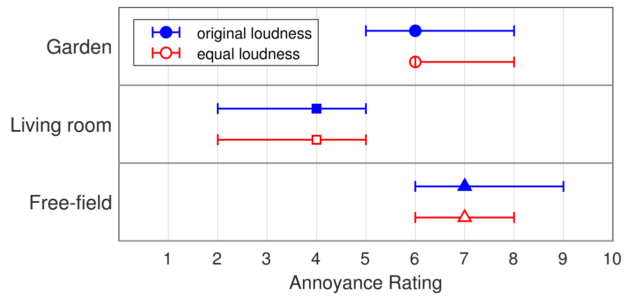


**Figure 4:** Median and interquartile ranges of the annoyance ratings for each scenario and loudness condition across all participants and stimuli. The *original loudness* condition is marked with blue filled symbols and the *equal loudness* condition is marked with red empty symbols.

| Loudness | Scenarios | $p$ | $r$ |
|---|---|---|---|
| Original | *Garden* vs *Living r.* | $< 0.001$ | 0.727 |
| | *Garden* vs *Free-f.* | $< 0.001$ | 0.329 |
| | *Living r.* vs *Free-f.* | $< 0.001$ | 0.793 |
| Equal | *Garden* vs *Living r.* | $< 0.001$ | 0.806 |
| | *Garden* vs *Free-f.* | $< 0.001$ | 0.414 |
| | *Living r.* vs *Free-f.* | $< 0.001$ | 0.840 |
| Original vs Equal | *Garden* | 0.903 | 0.004 |
| | *Living room* | 0.022 | 0.068 |
| | *Free-field* | 0.097 | 0.049 |

**Table 2:** Results of the Wilcoxon signed-rank tests comparing the scenarios within the two loudness conditions as well as between loudness conditions within each scenario. The $p$-values and effect sizes $r$ are given.



contrast, the comparison of the same scenarios in both loudness conditions revealed largely consistent ratings. Although a statistically significant difference was observed for the *living room* scenario ($p < 0.05$), the effect size was very small ($r < 0.1$), indicating no practically relevant difference.

A more detailed plot of the annoyance ratings is shown in Fig. 5. Depicted are the median and interquartile ranges of the annoyance ratings for all stimuli across participants in the *garden* (circles) and *living room* (squares) scenarios and for both loudness conditions (blue: *original loudness*, red: *equal loudness*). The stimuli of the *living room* scenario have median annoyance ratings between 3 and 4, with IQRs between 2 and 4 rating points, except for stimulus 26 at *original loudness*, which has an IQR of 5. The median annoyance ratings of the stimuli in the *garden* scenario are between 6 and 7, except for stimulus 9 in the *original loudness* condition, with a median annoyance rating of 5.5. The IQRs in the *garden* scenario are between 1 and 4, where 1 was only reached in the *equal loudness* condition and 4 only in the *original loudness* condition. Overall, in both scenarios, the IQRs are often wider in the *original loudness* condition than in the *equal loudness* condition. Between scenarios, the IQRs are more often wider in the *living room* scenario than in the *garden* scenario. Looking again at the median annoyance ratings, in both scenarios, the median rating between the loudness conditions is identical in 12 cases and differs in 15 cases. In cases of different median ratings, the deviation varies: for some stimuli, the median rating at *original loudness* is higher than at *equal loudness*, and for some stimuli vice versa. Also, per stimulus but in different scenarios, the median ratings do not always shift in the same direction, see, e.g., stimulus 5, where in the *living room* scenario the median rating at *equal loudness* is higher, but in the *garden* scenario the median rating at *original loudness* is higher. The different operating states of the heat pump (stimuli 1-22: *high-power heating* and stimuli 23-26: *maximum capacity*) do not seem to influence the annoyance ratings as no direct changes in the median ratings or IQRs can be identified.

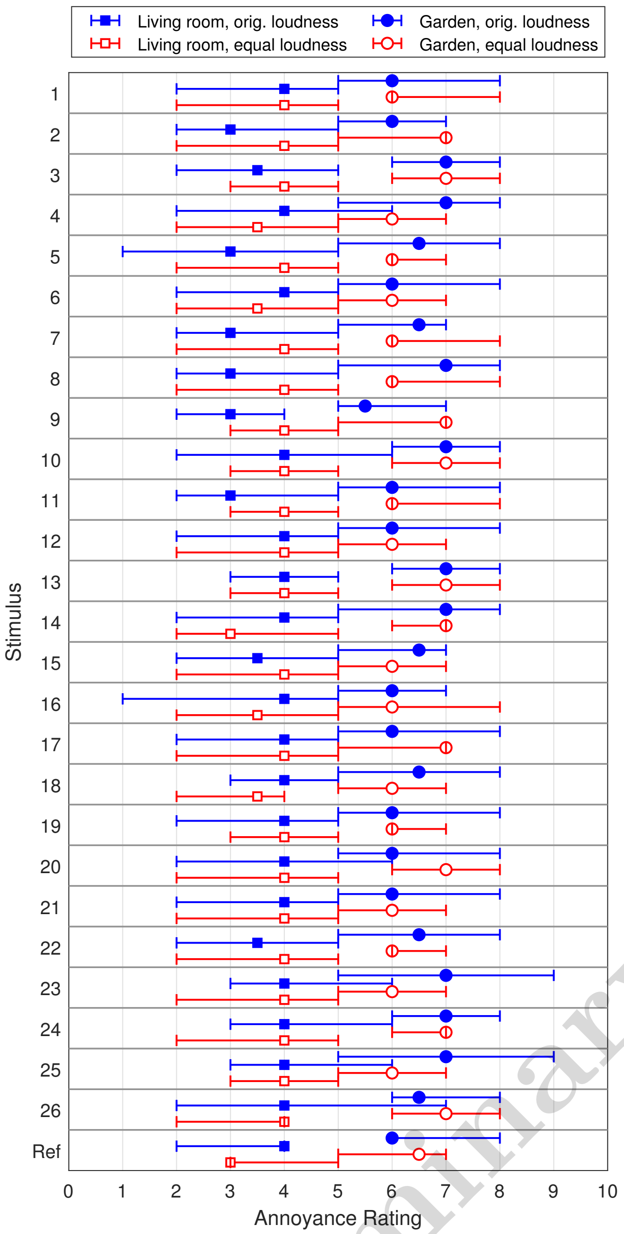


**Figure 5:** Median and interquartile ranges of the annoyance ratings for each stimulus. Direct comparison between the *garden* scenario (circles) and the *living room* scenario (squares) with *original loudness* (blue, filled symbols) as well as with *equal loudness* (red, empty symbols).

### 3.3. Own vs. neighbour's heat pump

Fig. 6 shows the differences in the annoyance rating between the two groups of participants that were instructed with the context that the sounds originate either from their *own heat pump* or from their *neighbour's heat pump*. A look at the median and IQRs in Fig. 6 shows that the median ratings are the same or at most one rating point different, and the IQRs also differ only minimally or are equal. Especially in the *living room* scenario, there is no difference between *own heat pump* and *neighbour's heat pump* in both loudness conditions. It is notable that the group *neighbour's heat pump* tended to rate the stimuli higher in annoyance, especially in the scenarios *garden* and *free-field* at *original loudness* as well as in the scenario *free-field* at *equal loudness*. Differences between

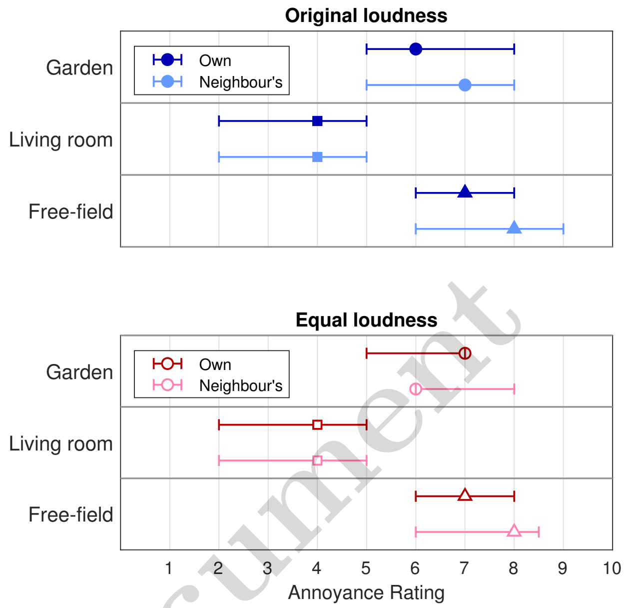


**Figure 6:** Median and interquartile ranges of the annoyance ratings for each scenario and both loudness conditions (top: *original loudness*, bottom: *equal loudness*) across all stimuli split into two participant groups referring to their *own* (dark colour) or to their *neighbour's* (light colour) heat pump.

the two participant groups are analysed using the Mann-Whitney $U$ test. The statistical significance is assessed based on the $p$-values, and effect sizes $r$ were also calculated to evaluate the practical relevance of the observed differences. The results of the analysis are shown in Table 3. In the scenarios *garden* at *original loudness* and *free-field* at both loudness conditions, the observed group differences are statistically significant ($p < 0.05$), but the results show very small effect sizes ($r < 0.1$), indicating only minor practical differences in the annoyance ratings.

| Loudness | Scenario | $p$ | $r$ |
|---|---|---|---|
| | *Garden* | 0.008 | 0.078 |
| *Original* | *Living Room* | 0.533 | 0.019 |
| | *Free-field* | 0.005 | 0.084 |
| | *Garden* | 0.148 | 0.043 |
| *Equal* | *Living Room* | 0.953 | 0.002 |
| | *Free-field* | 0.001 | 0.095 |

**Table 3:** Results of the Mann-Whitney $U$ tests for the two participant groups *own heat pump* and *neighbour's heat pump* in the different scenarios and loudness conditions. The $p$-values and effect sizes $r$ are given.

## 4. Discussions & Conclusion

The aim of this study was to investigate the influence of listener position, contextual framing, and loudness

on short-term annoyance of heat pump noise. Auralisation was used to simulate different sound propagation scenarios for heat pump sounds in residential areas, outdoors in a garden, and inside a living room. Additionally, as a reference, the sound propagation was simulated for a free-field environment. In a listening experiment, participants rated the annoyance of the stimuli across the three scenarios (*garden*, *living room*, *free-field*) using an 11-point rating scale. To investigate the impact of loudness on the annoyance rating, the stimuli were prepared not only at their original loudness but also at equal loudness, yielding two loudness conditions: *original loudness* and *equal loudness*. The reliability of the ratings was verified by repetition, resulting in a final dataset of 21 participants. The results show that the different scenarios lead to different annoyance ratings. The lowest annoyance ratings are reached in the *living room* scenario, the *garden* scenario was rated two points higher in the median, and the *free-field* scenario was rated one point higher in the median than the *garden* scenario. Looking at the mean loudness of the stimuli in the three scenarios (cf. Table 1), it seems that loudness is the main factor influencing the annoyance ratings. An influence on the ratings when setting the stimuli to the same loudness, in comparison to their original loudness, is not present. In detail, the rating differences between stimuli are at most 1.5 rating points, even though the heat pump stimuli differed in their sound characteristics and are distinguishable from each other. It seems that these characteristics do not influence the annoyance rating crucially. Dividing the participants into two groups and instructing them to rate the stimuli as if they are from their *own heat pump* or their *neighbour's heat pump* also does not lead to any major changes in the annoyance ratings. Even though the median ratings change in four of the six cases (cf. Fig. 6), the effect sizes are very small, and therefore the changes are not interpretable.
Acun et al. [1] did not find differences in the annoyance ratings between operating conditions, which is consistent with the findings of this study, but deviates from the results of Stürenburg et al. [2], in which differences in the annoyance ratings between operating conditions were found. These differences in the results could be due to the distance between source and receiver, which is closer in [2] than in the study presented here.
In an in-depth experiment, at equal sound pressure level the preference of heat pump stimuli can best be explained by sharpness and roughness, and not by loudness [6], but in the more basic experiment of the study presented here, using a rating scale, the short-term annoyance was mostly dependent on the loudness of the stimuli, which was dominated by the different auralised scenarios.
In general, the study shows that in a living room with a tilted window, annoyance across all heat pump stimuli was significantly reduced. Additionally, it was shown that the different operating conditions of the heat pump in the scenarios presented in this study, represented by the different stimuli, have no effect on the annoyance ratings. Further, it was shown that the annoyance ratings are independent of the contextual framing, whether it was one's own heat pump or the neighbour's heat pump.
One limitation of this study is the limited variation of heat pump sounds, as it only includes sounds from one heat pump. Another is that only two real-world scenarios were auralised; a broader set of simulated scenarios, with, e.g., different housing configurations, might give more insights.

## 5. Acknowledgments

The authors thank the student assistant, Judith Ziob, for her support in conducting the listening experiment, and the student assistant, Abdurrahman Alakus, for his support in evaluating it. The research presented here was funded by the German Federal Ministry of Economics Affairs and Climate Action as part of the project *LowNoise - Integrale Betrachtung, Optimierung und methodische Bewertung von Luft-Wasser-Wärmepumpen zur Reduktion akustischer Emissionen* (Funding Reference: 03EN4020A).

## References

[1] V. Acun, L. Barton, T. Cox, *et al.*, "Evaluating human response to air source heat pump noise for sustainable domestic heating," *Proceedings of the Institute of Acoustics*, vol. 46, no. 2, 2024.

[2] L. Stürenburg, J. Fels, V. Acun, *et al.*, "Study on human response to the cumulative impact of air source heat pump noise," in *15th IEA Heat Pump Conference 2026 (HPC2026)*, vol. Forthcoming article in Conference Proceedings, 2026.

[3] L. Stürenburg, H. Braren, L. Aspöck, and J. Fels, *Recordings of an Air-to-Water Heat Pump*, Zenodo, Sep. 2024. DOI: 10.5281/zenodo.13365535.

[4] C. H. Kasess, C. Reichl, H. Waubke, and P. Majdak, "Perception rating of the acoustic emissions of heat pumps," in *Forum Acusticum*, 2020, pp. 2453–2458.

[5] S. Sangsinsorn and B. Nienborg, "Noise immissions by air source heat pumps: A case study in Germany," *Building and Environment*, 2025.

[6] L. Stürenburg, H. Braren, L. Aspöck, and J. Fels, "Heat pump noise: Determination and modelling of preference-equivalent levels," *Acta Acustica*, vol. Forthcoming article, 2026. DOI: https://doi.org/10.1051/aacus/2026016.

[7] D. Schröder, *Physically based real-time auralization of interactive virtual environments*. Logos Verlag Berlin GmbH, 2011, vol. 11.

[8] Institute for Hearing Technology and Acoustics, RWTH Aachen University. "Room Acoustics for Virtual ENvironments (RAVEN) - a GA-based room acoustic simulation environment." https://www.virtualacoustics.org/RAVEN. (2017).