forum acusticum 2026
8-12 September · Graz, Austria

IMPACT OF ACOUSTIC CHARACTER ON HUMAN RESPONSE TO HEAT PUMP NOISE

Katie Salter[1], Simone Graetzer[1], Antonio Torija Martinez[1]

[1] *Acoustics Innovation Institute, University of Salford, United Kingdom*

Abstract

Air Source Heat Pumps (ASHPs) represent an essential technology for the decarbonisation of home heating across Europe. As ASHP deployment becomes more widespread, the introduction of this new noise source may increase community noise disturbance. Current standards for assessing ASHP noise typically focus on the sound pressure level (SPL) and often do not consider other factors, such as acoustic character. Acoustic character describes how a sound is perceived, beyond the loudness. This study investigates the impact of acoustic character on human perception of ASHP noise, with a focus on tonality. Recordings of ASHPs and synthesised reference sounds were presented to participants in a paired comparison listening test. Participants chose their preferred sound, as the reference sound level was varied following an adaptive procedure. A preliminary data analysis shows that tonal loudness, the difference between C- and A-weighted SPL, tonality and loudness have a significant positive relationship with response. This finding indicates that inclusion of a tonal penalty in ASHP noise assessment standards may more realistically represent human perception.

Keywords: *heat pumps, psychoacoustics, tonality, noise*

1. Introduction

Air Source Heat Pumps (ASHPs) are essential for the decarbonisation of heating, and are being installed increasingly across Europe. This rising installation rate is introducing a new noise source with a distinct acoustic character that is tonal, has low frequency content and contains intermittent elements due to changes in operating mode and defrost cycles. Tonality describes the presence of distinctly identifiable frequencies, or narrow bands of frequencies, that may be audible as a whistle or hum. Low frequency noise may be perceived as a low rumble or “bass” sound. Such acoustic features affect how people perceive and respond to sound, so this relatively new noise source may increase community noise disturbance, particularly as existing noise policy and regulation may not assess ASHP noise appropriately.

Assessment of ASHP noise is often based wholly on the sound pressure level (SPL) at a defined location, for example at the property boundary, with only some national noise regulations including consideration of acoustic character [3, 4, 5]. This is in spite of the suggestion by the underpinning EU directive that acoustic character be considered [6]. In the UK, acoustic character is generally included in environmental noise assessments [7], but not in the basic ASHP noise assessments that system design and installation teams are required to complete [8].

Several studies have shown that perception of ASHP noise is not driven by loudness or SPL alone [9, 10]. A recent literature review [5] reported the psychoacoustic features of tonality, roughness, sharpness and fluctuation in the low frequencies have been found to influence perception, with tonality highlighted as a research priority. In a paired comparison test, [9] presented participants with pairs of ASHP and reference sounds, along with the questions “Which sound is louder?” and “Which sound do you prefer?”, to find the SPLs of equal loudness and preference. The influence of various psychoacoustic metrics on the difference between the SPLs of equal loudness and preference was evaluated. Tonality was found to best predict this difference as a standalone metric, although a combination of roughness and sharpness was identified as the overall best fit. In [10] tonality was again identified as an important metric, showing a significant correlation with annoyance response to ASHP stimuli. This study went beyond annoyance and found tonality also significantly predicted Arousal (which describes the intensity of emotional response) and Valence (which is associated with the perceived pleasantness). It is clear that tonality is an important acoustic

Corresponding author: *k.salter@edu.salford.ac.uk.*

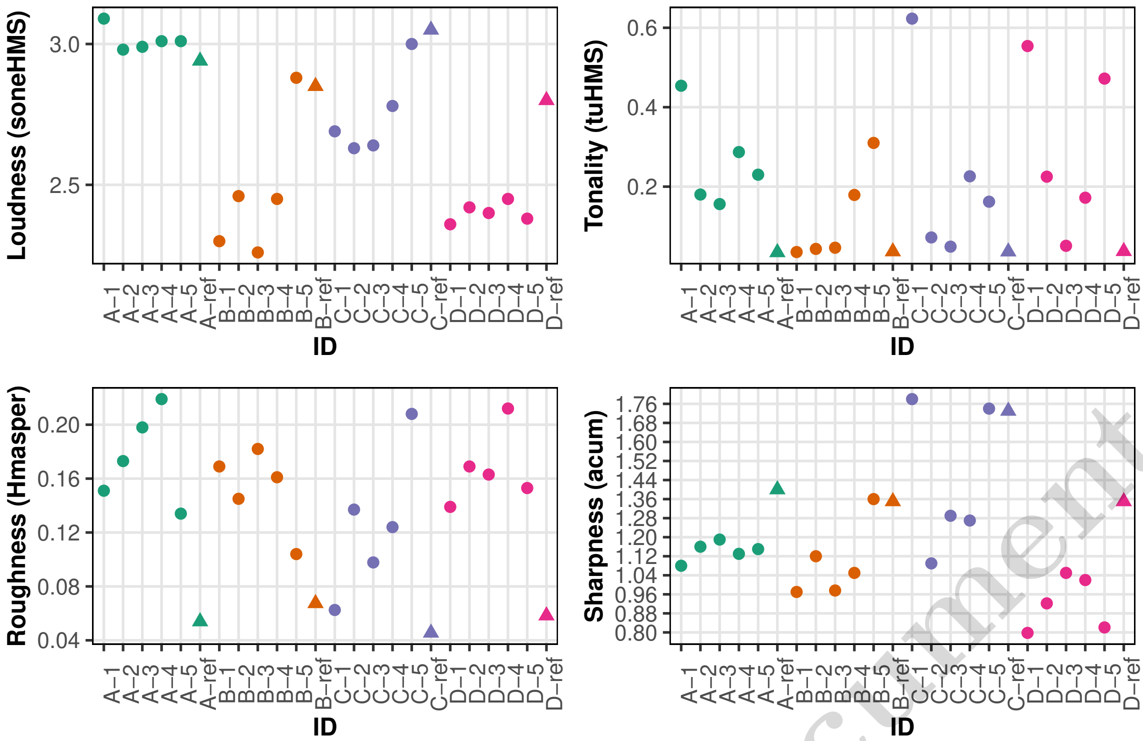

Figure 1: Loudness, tonality, roughness and sharpness of the ASHP sounds used in the experiment. The colours denote ASHP group: (a) green, (b) orange, (c) blue and (d) pink. Circles are ASHP sounds; triangles are reference sounds. The grid spacing on the y-axis indicates the Just Noticeable Difference: 0.5 sone for loudness 0.4 asper for roughness and 0.06 acum for sharpness [1, 2].

feature for annoyance response to ASHP sounds, but further evidence is needed to meaningfully incorporate an ASHP-specific tonality correction into noise assessments.

The present study focuses on tonality as the primary independent variable, with the variation in other psychoacoustic parameters restricted as far as possible. A paired comparison test, similar to the method of [9], has been used to identify the SPL at which ASHP sounds are judged to be subjectively equal in preference to non-tonal reference sounds. By including four ASHP units, with a particular reference sound simulated to be spectrally similar to each unit, but with no prominent tones, this study aims to investigate the impact of tonal content on preference of ASHP sounds. At the time of writing, data collection is ongoing, so here is presented a preliminary analysis of a subset of the data collected so far. The data in this preliminary study is from the first 20 study participants. The stimuli assessed by each participant have been pseudorandomised so that the number of auditions of each stimulus is fully balanced across every five participants. The pseudorandomised playlists heard by later participants and the information and instructions provided to participants during the study are fixed, so have not been influenced by this preliminary analysis.

2. Methods

2.1. Sound selection and preparation

2.1.1. ASHP sounds

Recordings from four different ASHP units were used. Recordings of an 11.2kW Mitsubishi Electric PUZ-WM112YAA ASHP were provided by the RWTH Aachen; see [11]. The recordings were taken using a 1/2 inch G.R.A.S 40HL low-noise microphone in a hemi-anechoic chamber. The microphone was placed 3m in front of the ASHP, off-axis to the fan airflow. The other three sets of ASHPs sounds were provided by the Fraunhofer Institute. These three ASHPs were of the same monobloc type and manufacturer as each other, but with different heating capacities: 7kW, 11kW and 12kW. The recordings were taken using a Head Acoustics HSU III.2 artificial head in a hemi-anechoic chamber. The Head acoustics recording process automatically equalises the recording to compensate for the spectral influence of the artificial head, so no further processing was required. The artificial head was placed 1.5m in front of the ASHP, off-axis to the fan airflow. For each sound a 3s perceptually stationary segment was selected, and a 5ms Hann fade was applied at the onset and end of the segment. The sounds used in the experiment are split by ASHP, which is referred to here as “group”.

The SPLs of all sounds at the listener position were normalised to 45dB(A). This was chosen as a realistic SPL that residents may be exposed to in a back garden, if one neighbour has an ASHP. The ASHP recorded at RWTH Aachen had an SPL of 45dB(A) at a distance of 3m when running in normal heating mode. In the UK the maximum allowed SPL from ASHPs is 37dB(A) at 1m from the facade of the neighbouring property [8] so a level of 45dB(A) in the garden is realistic. In the Netherlands the daytime noise limit for ASHPs is 45dB(A) at the property boundary [3]. Therefore, whilst the real-world SPL of ASHPs will depend on many factors, 45dB(A) is a level that could reasonably be expected to be commonly experienced by residents.

Various psychoacoustic metrics were calculated using ArtemiS Suite (HEAD acoustics GmbH, Germany). Loudness, tonality, roughness and tonal loudness were calculated using the Hearing Model of Sottek according to the ECMA-418-2 standard [12]. Sharpness was calculated using the DIN 45692 2009-08 (2009) standard [13]. In addition to this, the difference between the C-weighted and A-weighted sound levels (L_{C-A}) was calculated, which can indicate possible low frequency dominance. As the focus of this study was tonality, sounds were excluded if their value of sharpness, roughness, L_{C-A} or loudness was an outlier for that group, i.e. the value was outside of the the interquartile range (IQR) times 1.5 for that metric and group. Of the remaining sounds for each group, the descriptive statistics for the tonalities of the remaining sounds were calculated, and the sounds closest to the maximum, minimum, median, lower quartile and upper quartile values of tonality were chosen. Where two sounds were equidistant from one of these values, the sounds with the tonality value furthest from all other tonality values in that group was chosen. If this still did not lead to a single sound, the tonal loudness was then evaluated, and the sound that was furthest from all others in tonal loudness was selected. The final group of ASHP sounds comprised 20 sounds, with five sounds from each ASHP. Figure 1 shows the values of loudness, tonality, roughness and sharpness for the final 20 sounds that were selected.

2.1.2. Reference sounds

As the primary focus of this study is tonality, the reference sounds needed to be similar to the ASHP sounds, but with no tonal content. As each ASHP unit may have a slightly different sound character, a different reference sound was used for each of the four groups. The four references sounds were prepared by the following process for each group. For each sound, a peak finder and notch filter were used to remove up to 100 peaks in the spectrum with a prominence ratio of at least 6dB. The one-third octave band spectrum of each filtered ASHP sound was normalised such that the level of the 1kHz one-third octave bands were equal. The median one-third octave band spectrum for that group was then calculated, and applied as a filter to pink noise. As for the ASHP sounds, the sound were of 3s duration and included a 5ms Hann fade at the start and end.

Figure 2: Experiment set-up

2.2. Set-up and calibration

The test was conducted in the Listening Room at the University of Salford (UoS); this is a controlled laboratory environment. The sounds were played from MatLab 2025a, using a Motu 4Pre Audio interface, a Genelec 8030A loudspeaker and a Genelec 7050B subwoofer. The sounds were calibrated using a PCB 377B02 microphone with TEDS sensor, with a Dewesoft Sirius data acquisition system and an NTi Audio XL3 Class 1 sound level meter (SLM). Both microphones were calibrated using a Calibreuer BAC012 calibrator. Each sound was played and the gain adjusted until the $L_{Aeq,3s}$ matched the target level. In addition to the sounds used for the experiment, three pink noise signals at 35dB(A), 45dB(A) and 70dB(A) of 10s duration were calibrated as reference sounds for daily calibration. At the start of each day of testing the levels of the calibration reference sounds were checked using the SLM. If the level of the sounds varied by more than 0.5dB, the gain was adjusted using the audio interface. Figure 2 shows the set-up.

2.3. Experimental procedure

The experiment comprised a paired comparison between ASHP and reference sounds. The ASHP sound level was fixed 45dB(A), due to the differing perception of tonality at different sound levels [14]. The reference sounds were initially played at 45dB(A), and then varied following an adaptive staircase procedure. For each step, the participant heard the ASHP sound followed by the reference sound (or the two sounds in the reverse order, with the order randomised for each comparison), with a 0.5s gap. Participants had the option to play the pair of sounds a second time, with no further replays allowed after this. Participants were presented with the question: “You are in the garden doing a relaxing activity, for example reading, gardening or having a drink. You can hear one of these sounds continuously. Which sound do you prefer?”. This question and an image of a typical back garden were displayed throughout the playback. If the ASHP sound was preferred, the level of the reference sound was lowered. If the reference sound was preferred, the level of the reference sound was raised. This is based on the theory that louder sounds are generally

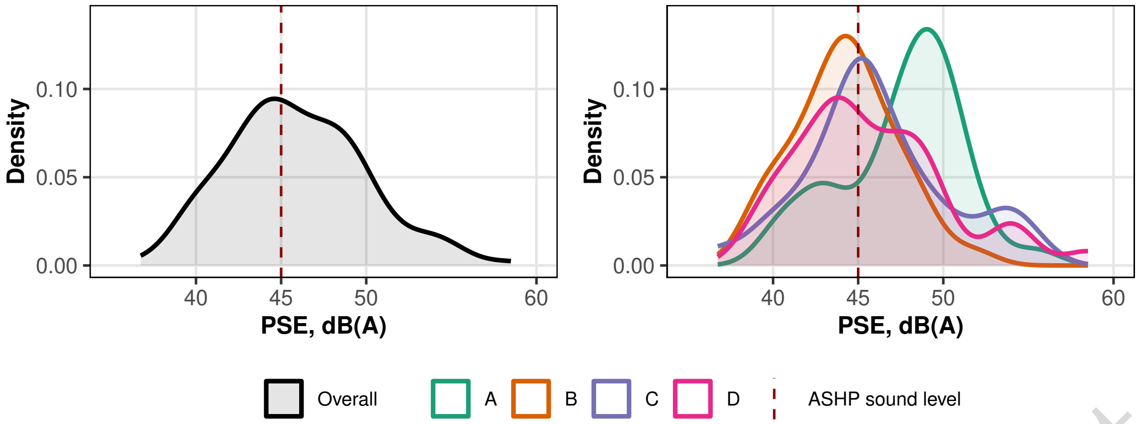

Figure 3: Probability density (kernel density) plot of PSE levels for all sounds (left) and split by stimulus group (right), calculated across all participants, with a density bandwidth of 1.4. The red dashed line shows the SPL of the ASHP sounds, which was fixed at 45dB(A).

less preferable, so lowering the reference sound level is expected to make it more preferable, and raising the level less preferable. The initial change in level was 6dB. If the participant changed their preferred sound from the ASHP to the reference sound (referred to as an "upward reversal"), the step size halved. The step size halved after each upward reversal to a limit of 1.5dB. The same pair of sounds was presented repeatedly, following this adaptive staircase procedure until a total of eight reversals (in either direction) was reached. The level of the reference sound that was considered to be the Point of Subjective Equality of preference (the PSE level) was then calculated as the mean of the levels preceding the last four reversals. The comparison cycle would also end after 20 comparisons, or if the reference sound level reached 25dB(A) or 70dB(A). In these cases, no result was calculated and the cycle was rejected. This procedure was based on previous similar studies [9, 15], to allow comparison of results. To limit the overall duration of the experiment, each participant listened to a playlist of eight sounds. The playlists were generated pseudorandomly, such that each playlist contained two sounds from each ASHP group, and the total number of auditions of each sound was balanced across participants.

3. Results and discussion

3.1. Participants

Participants were recruited through UoS and CIBSE[1] channels, LinkedIn and flyers posted around the university and local area. The participants included in this preliminary analysis are 30% female, 70% male; aged 18-24 (10%), 25-34 (55%), 35-44 (25%) and 45-54 (10%) years; 65% have not studied and/or worked in acoustics or audio; and the highest level of education completed is further education (10%), undergraduate degree (25%) and postgraduate degree (65%). All participants indicated that they were not aware of having any hearing loss or tinnitus.

[1] Chartered Institution of Building Services Engineers

3.2. Point of Subjective Equality of preference

Here is presented a preliminary analysis of the results from 20 participants. Within this sample each sound was auditioned a total of eight times, giving 160 data points. There were a total of 17 auditions that did not result in a final PSE level (here described as incomplete auditions), covering 12 different stimuli. Eight stimuli had an incomplete audition a single time, while four stimuli had an incomplete audition on two or three occasions. Seven of the incomplete auditions were assessed by a single participant; the others were assessed by seven different participants, each of whom had one or two incomplete auditions. 14 of the auditions were incomplete as 20 comparisons were reached before an eighth reversal occurred, and the remaining three reached the maximum allowed reference level. The participant who reached seven incomplete auditions has been excluded from further analysis. As they only reached one complete audition, it is not possible to review the distribution of this participant's results. After removing incomplete and excluded auditions, 142 auditions remain. 16 of the stimuli have seven or eight individual PSE results, and four have just five or six PSE results.

Figure 3 shows the probability density (kernel density) plots for the overall dataset and for each stimulus group. PSE levels above 45dB(A) indicate that the reference sound was generally preferred, as the reference sound reached a higher SPL before a participant changed their preference to the ASHP sound. Conversely, PSE levels below 45dB(A) indicate the ASHP sound was generally preferred, since the participants chose the ASHP sound even when at a higher SPL than the reference. Group A sounds are generally less preferable compared to their reference sound, whilst group B sounds are similar to or preferable compared

Table 1: Spearman's correlation coefficient and associated p-values between psychoacoustic metrics and PSE level.

Loudness (soneHMS)	Tonality (tuHMS)	Roughness (asper)	Sharpness (acum)	L_{C-A} (dB)	Tonal loudness (soneHMS)
.551*	.528*	.018 (*ns*)	.383 (*ns*)	.607**	.733***

ns = not significant, * $p < .05$, ** $p < .01$, *** $p < .001$

to their reference sounds. Groups C and D show a greater spread of results.

A Shapiro-Wilk test indicated that the distribution of PSE levels is not significantly different from normality, when considered as an overall dataset, at stimulus group-level and for individual participants, with one exception. For one participant, the distribution of PSE levels was found to be significantly different from normality (W = 0.79, p = 0.046), but with no significant skew or kurtosis. This participant had six complete auditions. The variances in PSE levels were similar between stimulus groups (with Levene's test F(3, 138) = 1.56, *ns*) and between participants (F(18, 123) = 1.34, *ns*). As mentioned above, one participant was excluded from further analysis due to having only a single complete audition. The tests of normality and homogeneity of variance were repeated with this participant included, with no change to the conclusions, therefore their exclusion does not significantly affect the overall distribution of results. The distributions of the psychoacoustic metrics shown in Figure 1 were significantly different from normality for all metrics except roughness, so the non-parametric Spearman's correlation rho was used to assess the relationship between PSE levels and psychoacoustic metrics.

Figure 4 presents the relationships between the median PSE level and tonal loudness and L_{C-A} , showing the PSE level increases with both of these metrics. Table 1 shows the Spearman's correlation results for loudness, tonality, roughness, sharpness, L_{C-A} and tonal loudness. Loudness, tonality and L_{C-A} are all significantly positively correlated with PSE level, with a moderate correlation. Tonal loudness shows a highly significant relationship with PSE level, with a moderate to strong correlation. Roughness and sharpness did not have a significant relationship with PSE level. This does not mean that roughness and sharpness do not influence perception in general – extreme values of roughness and sharpness were excluded.

These findings show that even when ASHP sounds are held at a fixed SPL, they are not perceived equally. Rather, other characteristics of the sounds affect human perception. Whilst the primary focus of this study was tonality, the complexity of and relationship between psychoacoustic metrics mean that variation in metrics other than tonality remained, albeit with extreme values removed. The correlation between L_{C-A} and PSE levels is noteworthy, as previous laboratory studies investigating human response to low frequency noise have had mixed findings [5]. It is important to note than L_{C-A} is not a definitive identifier of low frequency noise, rather an indicator of possible low frequency prominence, so other metrics for assessing low frequency noise will be considered in the full analysis.

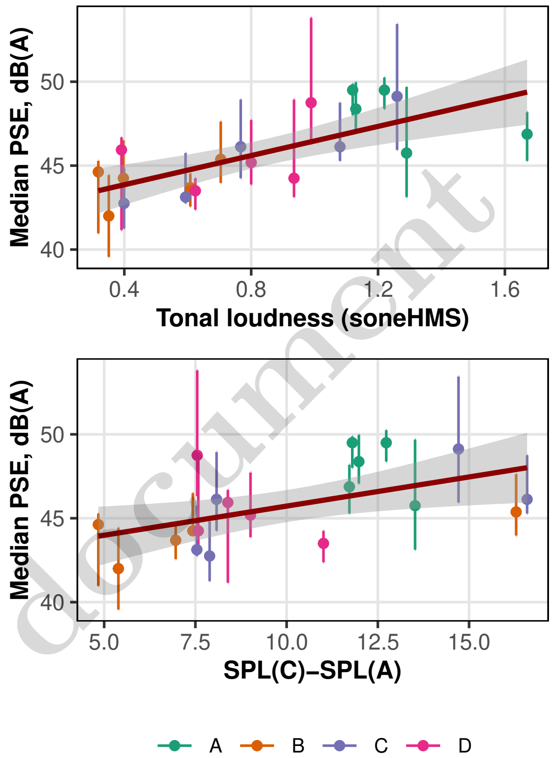

Figure 4: Relationship between median PSE level across all included participants and tonal loudness (upper) and L_{C-A} (lower). The whiskers indicate the IQR. The colours indicate the stimulus group.

4. Conclusions and further work

A preliminary analysis of results from a paired comparison listening study has been conducted. The study presented participants with recorded ASHP sounds and non-tonal synthesised reference sounds, varying the SPL of the reference sound until an end condition was met, allowing calculation of a PSE of preference between the ASHP and reference sounds. The PSE level had a significant positive correlation with tonal loudness, L_{C-A}, tonality and loudness. The results support the idea of adding a tonal penalty to ASHP noise assessment. Once data collection is complete, a full statistical analysis will be undertaken, considering various psychoacoustic metrics, including different metrics for tonality and low frequency noise.

5. Acknowledgements

Katie Salter would like to acknowledge the funding provided by the Engineering and Physical Sciences Research Council Centre for Doctoral Training on Sustainable Sound Futures (EP/Y034708/1) and CIBSE.

References

[1] J. You and J. Y. Jeon, "Just noticeable differences in sound quality metrics for refrigerator noise," *Noise Control Engineering Journal*, vol. 56, no. 6, p. 414, 2008, ISSN: 07362501. DOI: 10.3397/1.2987734. [Online]. Available: http://www.ingentaconnect.com/content/ince/ncej/2008/00000056/00000006/art00001 (visited on 03/27/2026).

[2] F. Pedrielli, E. Carletti, and C. Casazza, "Just noticeable differences of loudness and sharpness for earth moving machines," *The Journal of the Acoustical Society of America*, vol. 123, pp. 3164–3164, 5_Supplement May 1, 2008, ISSN: 0001-4966, 1520-8524. DOI: 10.1121/1.2933219. [Online]. Available: https://pubs.aip.org/jasa/article/123/5_Supplement/3164/616413/Just-noticeable-differences-of-loudness-and (visited on 03/27/2026).

[3] R. Fumagalli, P. Wagner, R. Pratter, *et al.*, "Annex 51: Acoustic Signatures of Heat Pumps, Final Report - Part 4, Deliverable 1.2 - Regulations - Countries overview," Heat Pump Centre, HPT-AN51-4, 2020. [Online]. Available: https://heatpumpingtechnologies.org/publications/final-report-part-4-annex-51-acoustic-signatures-of-heat-pumps/.

[4] R. Fumagalli, "Annex 51: Acoustic Signatures of Heat Pumps, Final Report - Part 5, Deliverable 1.3 - Regulations - Synthesis," Heat Pump Centre, HPT-AN51-5, 2022. [Online]. Available: https://heatpumpingtechnologies.org/publications/final-report-part-5-annex-51-acoustic-signatures-of-heat-pumps/.

[5] K. Salter, S. Graetzer, and A. J. Torija, "Review Of The Impact Of Acoustic And Non-Acoustic Factors On Human Response To Heat Pump Noise," in *Proceedings of the Institute of Acoustics*, vol. 47, Milton Keynes, 2025. [Online]. Available: https://www.ioa.org.uk/catalogue/paper/review-impact-acoustic-and-non-acoustic-factors-human-response-heat-pump-noise.

[6] *Directive 2002/49/EC of the European Parliament and of the Council of 25 June 2002 relating to the assessment and management of environmental noise - Declaration by the Commission in the Conciliation Committee on the Directive relating to the assessment and management of environmental noise*, 2002. [Online]. Available: https://eur-lex.europa.eu/eli/dir/2002/49/oj/eng.

[7] The British Standards Institution, *British Standard 4142:2014+A1:2019 Methods for rating and assessing industrial and commercial sound*, 2019.

[8] Microgeneration Certification Scheme, *MCS 020 a) Air Source Heat Pump Sound Calculation (For Permitted Development Installations)*, version ISSUE 1.0, 2025. [Online]. Available: https://mcscertified.com/wp-content/uploads/2025/03/MCS-020-a-Issue-1.0-Final.pdf.

[9] L. Stürenburg, H. Braren, L. Aspöck, and J. Fels, "Heat pump noise: Determination and modelling of preference-equivalent levels," *Acta Acustica*, vol. 10, p. 20, 2026, ISSN: 2681-4617. DOI: 10.1051/aacus/2026016. [Online]. Available: https://acta-acustica.edpsciences.org/10.1051/aacus/2026016 (visited on 04/15/2026).

[10] V. Acun, S. Graetzer, M. Radivan, and A. J. Torija, "Human response to air source heat pump noise: Influence of background noise, operating conditions and acoustic characteristics," *Acta Acustica*, vol. 10, p. 36, 2026, ISSN: 2681-4617. DOI: 10.1051/aacus/2026032. [Online]. Available: https://acta-acustica.edpsciences.org/articles/aacus/abs/2026/01/aacus250155/aacus250155.html (visited on 05/18/2026).

[11] L. Stürenburg, H. Braren, L. Aspöck, and J. Fels, "Recordings of an Air-to-Water Heat Pump," Sep. 12, 2024. [Online]. Available: https://zenodo.org/records/13365535 (visited on 05/16/2025).

[12] Ecma International, *ECMA-418-2 Psychoacoustic metrics for ITT equipment — Part 2 (methods for describing human perception based on the Sottek Hearing Model)*, version 4th edition, 2025. [Online]. Available: https://ecma-international.org/publications-and-standards/standards/ecma-418/.

[13] DIN, *DIN 45692:2009-08 Measurement technique for the simulation of the auditory sensation of sharpness (Messtechnische Simulation der Hörempfindung Schärfe)*, Aug. 2009. DOI: 10.31030/1521326. [Online]. Available: https://www.dinmedia.de/de/-/-/117635111 (visited on 08/01/2026).

[14] J. M. Francis, J. Lee, and L. M. Wang, "Determining Annoyance Thresholds of Tones in Noise," *ASHRAE Transactions*, vol. 120, 2014, ISSN: 00012505.

[15] S. Töpken and S. van de Par, "Determination of preference-equivalent levels for fan noise and their prediction by indices based on specific loudness patterns," *The Journal of the Acoustical Society of America*, vol. 145, no. 6, pp. 3399–3409, 2019, ISSN: 0001-4966. DOI: 10.1121/1.5110474.