PSYCHOACOUSTICAL INVESTIGATION OF SOUNDS FROM HEAT PUMPS

Carolin FELDMANN[1], Thomas CAROLUS[1],
Marc SCHNEIDER[2]

[1] *UNIVERSITY OF SIEGEN, Institute for Fluid- and Thermodynamics, Paul-Bonatz-Str. 9-11, 57076 Siegen, Germany*

[2] *EBM-PAPST MULFINGEN GMBH & CO. KG, Bachmühle 2, 74673 Mulfingen, Germany*

SUMMARY

Heat pumps become increasingly popular as heating systems for private households. The air delivering fan is an essential component in an air-to-air heat pump and simultaneously the main acoustic source. The overall aim of this work is to evaluate the quality of sound from heat pumps.

In a first step we are seeking a correlation of the overall perceived quality of the sound and more subjective descriptors like hissing, wobbling etc. The database is obtained by jury tests via the method of the semantic differential developed earlier specifically for sounds from fans and air delivering systems. In a second step the jury persons were interviewed individually. Eventually, we try to correlate the subjective descriptors like hissing, wobbling etc. with objective psychoacoustic metrics such as loudness, sharpness, roughness and fluctuation strength.

The interviewed participants favour heat pumps that sound dark or dull but not bellowing or humming. Whistling and hissing are rated as negative. The pitch of the sound should not be too high. The participants did not necessarily feel annoyed by the presence of tones, as long as they were reasonably monotonous in time and frequency. Undesired time structure effects are rattling, clattering or fluttering. The objective psychoacoustic metric sharpness turned out to be an important indicator for the sound quality. Disappointingly, the relevant time structure of the sounds is not reflected by any other objective psychoacoustic metric investigated. Here, further research is required. Nevertheless, at least 55 % of the variance in the subjective sound quality rating can be explained by the sharpness.

INTRODUCTION

For designing a technical system the sound quality can be of eminent importance. Heat pumps become increasingly popular as heating systems for private households. Purchasing a heat pump or replacing the conventional heating system in favor of a heat pump is mainly driven by a growing ecological sensibility, costs for energy and legal requirements. The air delivering fan is an essential component in an air-to-air heat pump and simultaneously the main acoustic source. Typically, air-to-air heat pumps are placed outside of the house, and hence a source of annoyance for the neighborhood. According to the German regulations the immission level in residential areas should not exceed 55 dB(A) by day and 40 dB(A) by night [1]. The corresponding rating level is defined as

$$L_r = L_{Aeq} + K_I + K_T + K_R + K_S \qquad \text{dB(A)} \tag{1}$$

with penalties for the impulsiveness K_I, tonality K_T, the count for periods of rest K_R and designated situations K_S [2]. Although this metric contains already A-weighting and the mentioned different penalties, the subjective perception of sounds of various heat pumps with equal rating levels is often completely different.
While different metrics exist for a number of hearing sensations such as loudness, sharpness, roughness etc., there is no general metric for the sound quality of arbitrary technical systems. As an example, based on the common psychoacoustic metrics, ALTINSOY [3] proposed two different models for the degree of annoyance of sound from vacuum cleaners and dish washers. Basically they are a linear combination of the common objective psychoacoustic metrics.

Our overall aim is to evaluate the quality of sound from the technical system 'heat pump', but with some more intermediate methodological steps as e.g. in ALTINSOY. In a first step we are seeking a correlation of the overall perceived quality of the sound and more subjective descriptors like hissing, wobbling etc. The database is obtained by jury tests via the method of the semantic differential. The semantic differential used had been developed earlier specifically for sounds from fans and air delivering systems [4]. In a second step the plausibility of the correlations obtained by the statistical analysis of the database is checked by interviewing the jury persons individually. Eventually, we try to correlate the subjective descriptors like hissing, wobbling etc. with objective psychoacoustic metrics such as loudness (DIN 45631/A1, ISO 532-1 [5, 6]), sharpness (DIN 45692 [7]), roughness and fluctuation strength [8, 9] of the sounds considered. An ultimate model which predicts the degree of annoyance from heat pumps remains a task for the future.

METHOD

Selection of sounds

28 recorded sounds from heat pumps were chosen from a huge database, cp. Table 1. The sounds represent air-to-air heat pumps and had been measured in an anechoic chamber. An artificial broadband noise, a pink noise filtered at 1.2 kHz so that it is comparable to the database sounds, was added as reference sound. No power-on or power-off sequences were considered for the sound sample. As the effect of loudness on perceived sound quality could hide all other influencing factors the range of the objective loudness is reduced by artificially equalizing all signals to an A-weighted sound pressure level of 55 dB(A).

Then the resulting loudness N of the 29 sounds ranges from 8.3 to 11.4 sone (DIN 45 631/A1) and the sharpness S (DIN 45 692), a measure for the amount of high frequency components in a sound, from 0.7 to 1.6 acum. The range of more objective psychoacoustic metrics of the sounds investigated are compiled in Table 1.

Table 1: List of sounds from heat pumps selected for the jury test

Sound No.	Heat pump No.	Fan type / Ø / rotor speed -/mm/rpm	*BPF* Hz	L_p(A) dB(A)	Objective sound metrics *N* sone	*S* acum	*R* c-asper	*F* c-vacil	*N*5/*N*95 -
1	1	Axial / 254 / 1188	99	55	10.3	1.2	8.1	0.9	1.10
2	1	Axial / 300 / 849	71	55	9.6	1.0	8.9	0.9	1.11
3	2	Axial / 496 / 971	81	55	10.6	1.6	7.3	0.7	1.09
4	2	Axial / 496 / 403	34	55	9.3	1.2	7.9	1.3	1.11
5	2	Axial / 496 / 558	47	55	9.4	1.2	8.1	0.9	1.10
6	3	Axial / 500 / 653	54	55	8.5	0.7	7.6	1.1	1.11
7	3	Axial / 500 / 646	54	55	9.1	0.8	6.6	1.3	1.13
8	3	Axial / 500 / 654	55	55	9.2	0.8	7.0	1.2	1.12
9	4	Axial / 626 / 902	75	55	10.1	1.4	7.4	0.8	1.08
10	4	Axial / 626 / 903	75	55	10.3	1.1	8.6	1.0	1.11
11	5	Axial / 626 / 504	42	55	9.4	0.9	8.9	0.9	1.11
12	5	Axial / 626 / 503	42	55	8.3	0.9	9.7	1.2	1.15
13	6	Axial / 703 / 909	76	55	10.5	1.3	8.1	1.0	1.14
14	6	Axial / 703 / 688	80	55	10.6	1.3	7.3	0.9	1.11
15	6	Axial / 703 / 665	78	55	10.6	1.4	7.0	0.9	1.11
16	6	Axial / 703 / 668	78	55	11.4	1.1	6.9	1.0	1.17
17	6	Axial / 703 / 685	80	55	11.1	1.3	7.4	0.9	1.11
18	6	Axial / 703 / 533	62	55	10.7	1.1	7.6	0.9	1.13
19	7	Axial / 703 / 907	76	55	9.7	1.3	8.0	0.8	1.09
20	7	Axial / 703 / 461	38	55	9.0	1.1	9.1	1.0	1.10
21	6	Axial / 800 / 933	78	55	10.8	1.3	7.7	0.9	1.12
22	6	Axial / 800 / 835	70	55	10.4	1.3	7.8	0.9	1.12
23	8	Radial / 250 / 1979	231	55	9.8	1.1	8.0	0.9	1.09
24	9	Radial / 450 / 979	98	55	10.2	0.8	7.1	1.0	1.15
25	9	Radial / 450 / 978	98	55	9.5	0.7	6.6	1.4	1.22
26	10	Radial / 500 / 800	80	55	11.3	1.3	6.7	0.9	1.12
27	10	Radial / 500 / 584	58	55	10.9	1.2	7.5	0.9	1.12
28	4	Radial / 630 / 942	94	55	10.4	1.1	7.6	1.0	1.11
29		synthetic noise		55	10.4	1.4	8.0	0.8	1.09

BPF = blade passing frequency, i.e. number of rotor blades x rotational speed of rotor, loudness N accord. to DIN 45 631/A1, sharpness S accord. to DIN 45692, roughness R accord. to [8], fluctuation strength F accord. to [9]

Jury test - jury and semantic differential

The jury comprised 19 male and 21 female participants (median age 23 years). The jury test was carried out in an acoustically optimized room. A maximum of six participants were assessing the sounds simultaneously via Sennheiser HD 650 headphones and the playback system PEQ V®

(HEAD acoustics). The test was implemented with the Software SQuare® (HEAD acoustics). Before the test started all participants were briefed in written form. After briefing and prior to the main test all sounds were presented for 5 s each to the participants to provide a first survey of the stimuli. When the participants felt confident the test started. The jury test is based on a semantic differential, which was established in a previous study specifically for fan related sounds [4]. The semantic space used consists of 37 different adjective pairs, Table 2.

Table 2: List of bipolar and unipolar adjective scales used in the jury test

bipolar scales	
1	loud - soft (laut - leise)
2	aggressive - relaxed (aggressiv - entspannt)
3	intrusive - unintrusive (aufdringlich - unaufdringlich)
4	unpleasant - pleasant (unangenehm - angenehm)
5	muffled - shrill (dumpf - schrill)
6	dark - light (dunkel - hell)
7	dark - light (dunkel - hell) [REPETITION]
8	dull - sharp (stumpf - scharf)
9	low - high (tief - hoch)
10	irregular - regular (ungleichmäßig - gleichmäßig)
11	agitated - calm (unruhig - ruhig)
12	animated - static (bewegt - statisch)
13	rough - smooth (rau - glatt)
14	unsteady - steady (instationär - stationär)
15	slow - fast (langsam - schnell)
16	low speed - high speed (niedertourig - hochtourig)
17	heavy - light (schwer - leicht)
18	weak - strong (kraftlos - kräftig)
19	small - large (klein - groß)
20	weak - strong (schwach - stark)
21	powerless - powerful (leistungsschwach - leistungsstark)
22	cheap - high class (billig - hochwertig)
23	coarse - smooth (grob - sanft)
24	hard - soft (hart - weich)
25	hollow - solid (hohl - massiv)
unipolar scales, ranging from "completely" to "not at all"	
26	annoying (lästig)
27	bothersome (störend)
28	possible to fade out (ausblendbar)
29	possible to fade out (ausblendbar) [REPETITION]
30	humming (brummend)
31	droning (dröhnend)
32	bellowing (röhrend)
33	whistling (pfeifend)
34	abrasive (schleifend)
35	hissing (zischend)
36	fluctuating (fluktuierend)
37	wobbling (schwankend)
38	noisy (rauschhaft)
39	tonal (tonhaltig)

English translation slightly modified as compared to [4];
REPETITION = multiple occurrence for reliability check - see appendix

The jury test was subdivided in two parts with 15 sounds each (V_1 and V_2). The sounds for each part were chosen in the way that the sharpness range as well as the roughness range is similar within both parts. As the roughness has only a small variance and not all possible time structure effects are

measured with this metric, the final decision was assisted by the own perception. The participants group was split in two groups as well. The first group P_{12} started with the first part of the test V_1. The second test V_2 was done with at least one day of recovery. The second participant group P_{21} started in reverse order. The evaluation part is separated in three parts ($V_{x,1}$, $V_{x,2}$ and $V_{x,3}$). In every part the participants had to evaluate all the sounds on 13 adjective scales.

Interview

Subsequent to the jury test an interview session was carried out. The interview was held in written form. First of all the jury test persons were asked, if the device 'heat pump' was known before the jury test started. Moreover, they were requested to describe the character of a pleasant or unpleasant heat pump sound. Finally the participants had to imagine being in a garden or a balcony and would hear the heat pump sounds of the listening test. They were asked then if heat pump sound in general is acceptable or not. The answers were used to suggest assumptions for the analysis of the semantic differential.

RESULTS

Subjective evaluation of heat pump sounds

Interview. Only 44 % knew the device heat pump before they were instructed for the jury test. Although almost half of the jury did not know heat pumps the opinion what characterizes a pleasant sound was unambiguous. The participants favour heat pumps that sound dark or dull but not bellowing or humming. To quote test participants' rating: (i) "no whistling nor hissing, the pitch should not be too high", (ii) "the sound must not be too 'fast' but 'slow'". Moreover, the participants did not necessarily feel annoyed by the presence of tones, as long as they were reasonably 'monotonous' in time and frequency.

Much more complains were associated with broadband noise. Undesired time structure effects are 'rattling', 'clattering' or 'fluttering'. The participants answered that a heat pump is accepted in the garden if the sound has the described positive characteristics and the heat pump is not placed too close to the listener.

Semantic differential. The resulting data matrix (40 participants x 30 sounds x 39 adjective scales) is reduced using two principal component analysis (PCA) by carrying out the smallest number of explanatory constructs named components that explain a maximum of common variance in a correlation matrix of the adjective scales or the sound files, cp. e.g. FIELD [10], BORTZ and DOERING [11] or BUEHNER [12]. Five components *spectral content 1 (SC1)*, *quality (Q)*, *time structure (TS)*, *spectral content 2* (SC2) and *power (P)* are constructed by the first PCA, reducing the amount of subjective descriptors (adjective scales). The adjective scales for describing spectral differences are subdivided in two components. Describing bipolar adjective pairs like "dark - light" are categorized in the *SC1*. Onomatopoeic adjective pairs, which are rated on a unipolar scale like "humming - not humming" or "whistling - not whistling" are separated by their spectral centre. Only adjective pairs which explain high frequency effects, e.g hissing, are categorized in *SC2*.

The second PCA, that reduces the amount of sounds, yields four different sound groups. The results of both PCA's are depicted in the polar diagram in Figure 1. The evaluations of the sounds are averaged over their sound group and also over all sounds ("mean"). The first group of sounds found in the PCA represents the lowest evaluation values in *SC1*, cp. Figure 1. Sound group 2 clusters signals which are light, cp. *SC1*, but not very unpleasant (*Q*) and as well not very time varying (TS). The sounds from sound group 3 seem to be the most annoying. Although they are similar on the

scale "dark - light" to sound group 2, they are much more unpleasant because of the perception in TS and *SC2*.

It can also be learned from Figure 1 that although sound group 2 and 4 are close together on many scales of *SC1* they are differed in *SC2* in the way that sound group 4 contains sounds that are light but not hissing or whistling.

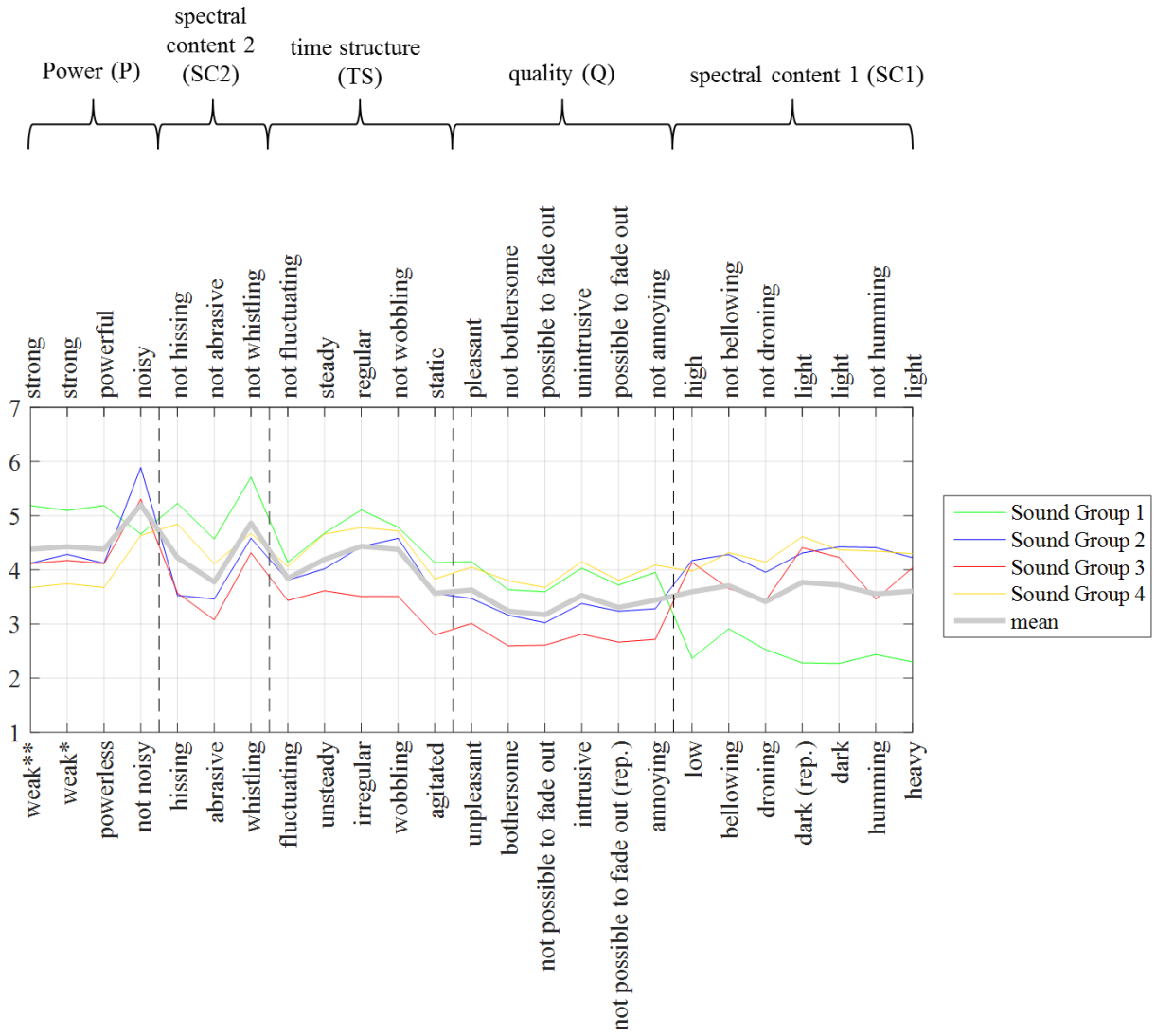

Figure 1: Polar diagram of jury test results. The adjectives are sorted by the components as obtained by the principal component analysis (PCA). The sound groups extracted from the second PCA are represented by their mean values for all adjective scales. The scale "weak - strong" is marked with (*) as translation of the original scale "schwach - stark" and marked with (**) as translation of the original scale "kraftlos - kräftig"

Now we are seeking a correlation of the overall perceived quality *Q* of the sounds and the more subjective descriptors collected in *SC1*, *TS* and *SC2*. As the power *P* just explains 6 % of the variance in the PCA, we do not expect a high impact on quality from this component and therefore exclude it from further analysis. In Figure 2 the perceived quality *Q* is plotted as a function of the subjective descriptors *SC1*, *TS* and *SC2*. The first diagram (Figure 2 left) shows the relation of *Q* and *SC1*. It is important to remember that *SC1* predominantly reflects the sensation of a 'dark', 'humming', 'drowning' etc. sound and *SC2* the sensation of 'hissing' and 'whistling'; the scales of *TS* and *SC1/2* range from "completely perceived" = 1 to "perceived not at all" = 7. Obviously, *Q* is not depending on *SC1*. The middle and right plot in Figure 2, however, suggest a linear correlation of *Q* with *TS* and *SC2*, the latter showing the highest correlation coefficient $r = 0.66$.

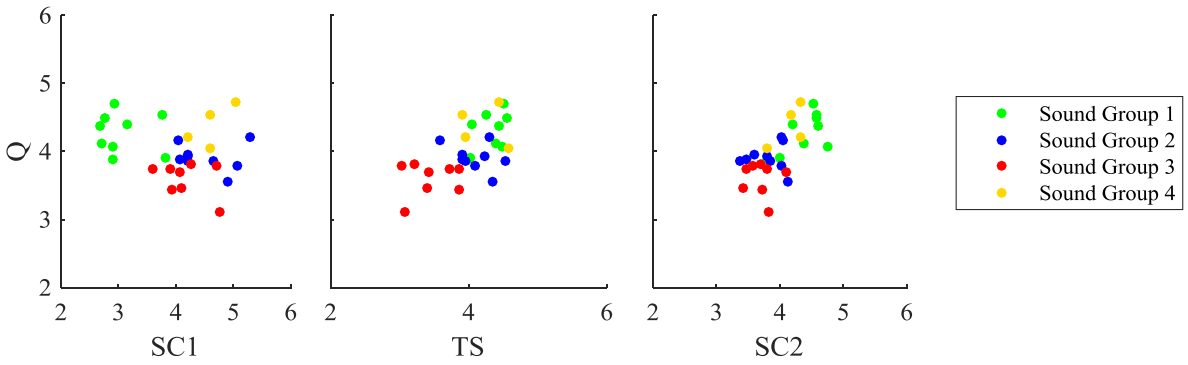

Figure 2: Q as a function of SC1, TS, and SC2; the colors represent the different sound groups obtained from a principal component analysis (PCA)

This outcome is partly confirmed by the interviews, since

- The component time structure *TS* comprises 'fluctuating', 'unsteady', 'irregular', 'wobbling', 'agitated' and their opposing antonyms - this corresponds well to 'rattling, clattering and fluttering' and 'non-monotonous' in time and frequency from the interview;
- the component *SC2* comprises 'hissing', 'abrasive', 'whistling' and their opposing antonyms - these terms correspond well to the statements in the interview 'no whistling nor hissing, the pitch should not be too high'.

The outcome from the interview also suggests an effect of *SC1* (e.g. humming), which is not seen by the statistics in Figure 2 left.

Correlation to psychoacoustic metrics

To build up a sound quality metric for heat pumps we finally have to find objective metrics for the main factors of the quality perception. Common metrics associated with *TS* are the roughness [8] and the fluctuation strength [9], and associated with *SC2* the sharpness. We now correlate the objective metrics of all sounds in Table 1 with the values of the jury tests for *TS* and *SC2* as in Figure 2. Table 3 shows the results. The sharpness is a good descriptor of *SC2* with a highly significant correlation of r = -0.78.

Table 3: Pearson correlation between the principal components "spectral content II" and "time structure" and common psychoacoustic metrics sharpness, roughness and fluctuation strength.

	S acum	*R* asper	*F* vacil	N_5/N_{95} -
TS	-	r = -0.11	r = 0.08	r = 0.11
SC2	r = -0.78**	-	-	-

** outlines significant correlation coefficients on a 0.01 level

On the other hand, a common metric to describe *TS* could not be found as the correlations are too low and not significant. Furthermore, the values from the metrics roughness *R* and fluctuation strength *F* are below the threshold found for this sensation in [9], cp. Table 1. This would imply that

these sensations are not included in the perceived time structure for heat pumps or the common algorithms do not compute very satisfactory.

As an exemplary alternative objective metric we consider the loudness of time variant sounds according to DIN 45631/A1 [5]. The ratio N_5/N_{95}, where e.g. N_5 is the five percentile loudness, i.e. only 5 % of the loudness values over time are higher than N_5. If this ratio is higher than 1.1, the sound is called time variant. This ratio could help to find time variant sounds with regular and irregular modulation behaviour, i.e. distinct modulation frequencies measurable or not. As seen in Table 1 the variance for this ratio, however, turned out to be too small to be a relevant indicator for different time structures of the sounds.

The correlations from diagrams Figure 2 middle and right suggest a linear quality prediction model as

$$Q = k_1 TS + k_2 SC2 + \varepsilon\,. \tag{2}$$

k_1 and k_2 are constants obtained from the correlation, ε is the error. An ansatz for the model which eventually would allow predicting the sound quality of heat pumps with exclusively objective metrics is

$$Q = k_a\left(\text{objective metric for } TS\right) + k_b\left(\text{objective metric for } SC2\right) + \varepsilon \tag{3}$$

with other constants k_a and k_b. As mentioned above, in our case the sharpness S was identified as the objective metric for $SC2$. Unfortunately, no objective metric was found for TS. Despite this fact, at least 55 % of the variance of Q can be explained by S. The remaining 45 % are buried in the missing metric for TS and the error.

CONCLUSIONS

28 different heat pump sounds were evaluated in a jury test using the method of semantic differential. The sounds could be separated in 4 perception groups using a principal component analysis. The main influencing factors for the perceived quality are the sensations of 'humming', 'hissing', 'wobbling' or 'agitated' learned from the jury test or 'hissing', 'whistling', 'rattling', 'clattering' and 'fluttering' from the subsequent interview session. In contrast to the traditional rating level L_r with its number of penalty terms the objective psychoacoustic metric sharpness turned out to be an important indicator for the sound quality. Disappointingly, the relevant time structure of the sounds is not reflected by any other objective psychoacoustic metric investigated. Here, further research is required. Nevertheless, at least 55 % of the variance in the subjective sound quality rating can be explained by the sharpness.

BIBLIOGRAPHY

[1] Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit – *Sechste Allgemeine Verwaltungsvorschrift zum Bundes-Immissionsschutzgesetz (Technische Anleitung zum Schutz gegen Lärm - TA Lärm)*. Berlin, Germany, **1998**

[2] DIN Deutsches Institut für Normung E.V., DIN 45645-1 – *Determination of rating levels from measurement data - Part 1: Noise immission in the neighbourhood*. Beuth Verlag, Berlin, Germany, **1996**

[3] M.E. Altinsoy – *Towards an European Sound Label for Household Appliances: Psychoacoustical Aspects and Challenges*. Proceedings of 4th International Workshop on Perceptual Quality of Systems, Vienna, Austria, **2013**

[4] C. Feldmann, T. Carolus, M. Schneider – *A semantic differential for evaluating the sound quality of fan systems*. Proceedings of ASME Turbo Expo, **2017**

[5] DIN Deutsches Institut für Normung E.V., DIN 45631/A1 – *Calculation of loudness level and loudness from the sound spectrum - Zwicker method - Amendment 1: Calculation of the loudness of time-variant sound*. Beuth Verlag, Berlin, Germany, **2010**

[6] ISO International organization for standardization, ISO 532-1 – *Acoustics - Methods for calculating loudness - Part 1: Zwicker method*. Geneva, Switzerland, **2017**

[7] DIN Deutsches Institut für Normung E.V., DIN 45692 – *Measurement technique for the simulation of the auditory sensation of sharpness*. Beuth Verlag, Berlin, Germany, **2009**

[8] R. Sottek – *Modelle zur Signalverarbeitung im menschlichen Gehör*. PHD Dissertation, RWTH Aachen, **1993**

[9] H. Fastl, E. Zwicker – *Psychoacoustics: Facts and models*. 3rd edition, Springer Verlag, Berlin, Germany, **2006**

[10] A.P. Field – *Discovering statistics using IBM SPSS statistics.* 4th edition, Sage Publ., Los Angeles, USA, **2013**

[11] J. Bortz, N. Döring – *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. 4th edition, Springer-Medizin-Verlag, Heidelberg, Germany, **2006**

[12] M. Bühner – *Einführung in die Test- und Fragebogenkonstruktion.* 3rd edition, Pearson Studium, Munich, Germany, **2011**

[13] H.F. Kaiser – *Little Jiffy, M. IV.* Educational and psychological measurement, Vol. 34, No. 1, pp. 111-117, **1974**

ANNEXES

Reducing the amount of subjective descriptors using a principal component analysis

The suitability of the dataset was checked by the criterion of KAISER-MEYER-OLKIN KMO (0.88 "meritorious", [13]) and BARTLETT's test of sphericity ($\chi^2(300)=11809.9$, $p<0.001$, cp. [10, 12]).

Table A1: Principal Component Analysis for reducing the dataset and grouping adjective scales to Components. Component loadings smaller than 0.3 are not shown in the Table. The scale "weak - strong" is marked with (*) as translation of the original scale "schwach - stark" and marked with (**) as translation of the original scale "kraftlos - kräftig" KMO=0.88, $\chi^2(300)=11809.9$, $p<0.001$

		Component				
		I	II	III	IV	V
Spectral Content I	heavy - light	0.76				
	humming - not humming	0.75				
	dark - light	0.74				
	dark - light (Rep.)	0.73				
	droning - not droning	0.70				
	bellowing - not bellowing	0.68				
	low - high	0.61				
Quality	annoying - not annoying		0.84			
	possible to fade out - not possible to fade out (Rep.)		-0.82			
	intrusive - unintrusive		0.78			
	possible to fade out - not possible to fade out		-0.75			
	bothersome - not bothersome		0.73	0.33		
	unpleasant - pleasant		0.60		0.42	
Time structure	animated - static			0.73		
	wobbling - not wobbling			0.68		
	irregular - regular			0.56	0.36	
	unsteady - steady			0.52		
	fluctuating - not fluctuating			0.53		
Spectral Content II	whistling - not whistling				0.73	
	abrasive - not abrasive				0.67	
	hissing - not hissing				0.56	-0.44
Power	noisy - not noisy					-0.62
	powerless - powerful	-0.35			0.34	0.54
	weak - strong*	-0.45				0.46
	weak - strong**	-0.55				0.41
	variance explained	17.3	15.3	9.3	9.1	5.8

Reducing the amount of sound files using a principal component analysis

The dataset is suitable with a KMO of 0.93 ("marvelous", [13]) and BARTLETT's test of sphericity ($\chi^2(435)=13646.9$, $p<0.001$, cp. [10, 12]).

Table A2: Principal Component Analysis for reducing the dataset and grouping Sounds to Components. Component loadings smaller than 0.3 are not shown in the Table. KMO=0.93, $\chi^2(435)=13646.9$, $p<0.001$

	Sound No.#	Component 1	2	3	4
Sound Group 1	7	0.86			
	24	0.85			
	6	0.85			
	8	0.83			
	25	0.77			
	18	0.74			
	16	0.69		0.35	
	11	0.60			0.43
	10	0.37		0.34	
Sound Group 2	29 (rep.)		0.78		
	19		0.75		
	29		0.74		
	14		0.60		
	17		0.58		
	15		0.58	0.36	
	9		0.56		
	27		0.53		
	28		0.48		
Sound Group 3	13		0.31	0.70	
	5			0.67	
	1			0.66	
	21		0.35	0.62	
	3	-0.30	0.38	0.61	
	22		0.31	0.55	
	4		0.32	0.49	
	26	0.36	0.39	0.45	
Sound Group 4	12				0.65
	20		0.49		0.64
	2			0.36	0.54
	23		0.30		0.42
	variance explained	19.3	15.7	12.7	6.1

Reliability

To test the reliability of the jury test results two adjective scales and one sound was repeated. One descriptive adjective scale "dark - light" was chosen as well as one evaluative adjective scale "completely possible to fade out - not at all possible to fade out". The adjective scales were distributed over the three sub parts of the test, so that not two identical scales appear in one sub part and that not too many descriptive or evaluating adjective scales are in one sub part in relation to the other. As repeated sound the reference sound 29 was chosen as most unsuspicious sound. If the data is reliable the mean of the evaluation is assumed to be stable from test to retest. Therefore the difference in mean between test and retest will be investigated, as shown in Figure A1. No mean differs from test to retest more than one scale point. So the data could be stated as reliable.

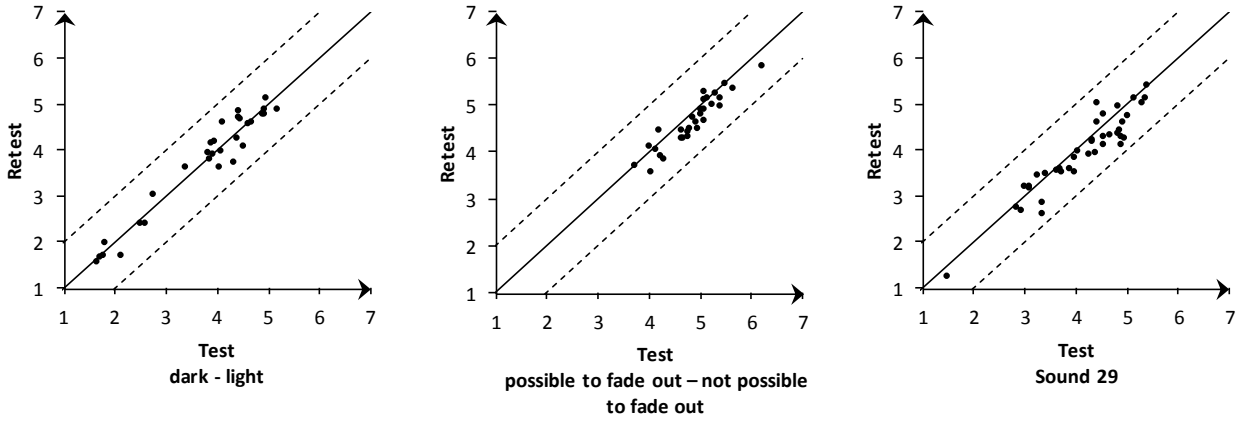

Figure A1: Mean evaluation for retest over test for the adjective scales "dark - light", "completely possible to fade out - not at al possible to fade outl", and for the sound 29 (from left to right)