

COVER PAGE

Document downloaded by @DAEL

Fri Nov 27 02:32:22 2026
For personal use

When automatic English translation is provided, only the original document is authentic.
The EAA cannot be held responsible of any translation error

Bibliographical reference

Psychoacoustic Characterization of the Noise Produced by Photovoltaic Inverters, Wolfgang Ellermeier, Florian Kattner, Lothar Kurtze and Joachim Bös, Acta Acustica **vol. 100** (Number 6), 2014, pp. 1120-1128

DOI
https://doi.org/10.3813/AAA.918791

ACTA ACUSTICA UNITED WITH ACUSTICA
Vol. 100 (2014) 1120 – 1128

DOI 10.3813/AAA.918791

Psychoacoustic Characterization of the Noise Produced by Photovoltaic Inverters

Wolfgang Ellermeier[1)], Florian Kattner[1)], Lothar Kurtze[2)*], Joachim Bös[2)]

[1)] Institut für Psychologie, Technische Universität Darmstadt, Alexanderstr. 10, 64283 Darmstadt, Germany.
ellermeier@psychologie.tu-darmstadt.de

[2)] Fachgebiet Systemzuverlässigkeit und Maschinenakustik SzM, Technische Universität Darmstadt

Summary

Photovoltaic inverters used in solar power systems produce noise across the entire audible spectrum. To assess the annoyance generated by such noise, listening tests were performed on binaural recordings made from different brands of inverters in various operating conditions. 42 participants judged how unpleasant each of 40 sound samples appeared when compared to a reference using the method of magnitude estimation. In addition, they performed all possible paired comparisons on a subset of the sounds. The paired-comparison data were largely consistent with the Bradley-Terry-Luce model implying that auditory unpleasantness of these sounds is ratio-scalable. The magnitude estimates were highly reliable, showed systematic dependencies on the type of solar inverter, and increased monotonically with the electrical power converted. They further indicated that the noise components due to cooling ventilation emerging in the mid-frequency range have a greater impact on perceived unpleasantness than the high-frequency peaks near 16–18 kHz. When trying to predict mean unpleasantness scale values on the basis of psychoacoustical metrics, a linear combination of (binaural) loudness, roughness, and sharpness accounted for over 90% of the variance in the subjective data. By contrast, A-weighted sound pressure level and other purely acoustical measures performed poorly in this respect. This indicates that it may be advantageous to include psychoacoustical parameters in the evaluation of noise generated by solar inverters.

PACS no. 43.66 Cb, 43.66 Yw, 43.50 Ba, 43.50 Jh

1. Introduction

When generating electricity from solar energy, the direct current produced by photovoltaic panels must be converted to alternating current to be fed back to the power grid. To accomplish this, *photovoltaic inverters* are used in virtually every solar power system. Due to the electronic and mechanical components built into them, these inverters produce noise nearly across the entire audible frequency spectrum.

Specifically, the power electronics used in the inverters – thyristors and inductance coils – often generates high-frequency tonal components while the ventilation used for cooling might add broadband noise in the low to mid-frequency range. Figure 1 illustrates the type of spectrum generated by a photovoltaic inverter – one of the devices investigated in the present study – with a prominent peak at 18 kHz and a further harmonic at twice that frequency.

Since photovoltaic inverters are typically mounted inside of human dwellings, e.g., in corridors or basements, customer complaints about annoying sound emissions have become an issue, and the industry has begun to be concerned about the effects of inverter noise on humans and animals. To our knowledge, however, no systematic study of the perceptual effects of the noise produced by photovoltaic inverters has been performed, and there are presently no standards or regulations with respect to the acoustic properties of the devices. In any event, it appears unlikely that a simple measure such as A-weighted sound pressure level will suffice to characterize their auditory effects.

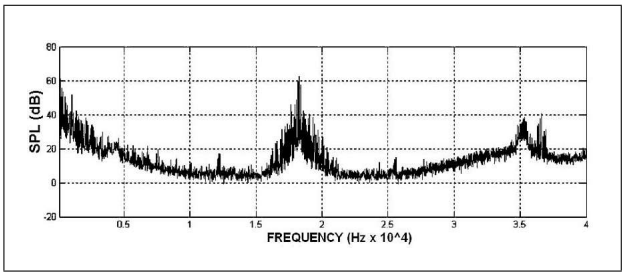

Figure 1. Sound pressure spectrum of inverter A as measured using a half-inch ultrasound microphone.

While this may be of concern to applied acoustics, inverter sound also constitutes an interesting stimulus from a basic psychoacoustic perspective: It is a relatively homogeneous set of tone-noise mixtures that lends itself to

Received 21 August 2013,
accepted 26 May 2014.

* presently with Geislinger GmbH, Salzburg, Austria

© S. Hirzel Verlag · EAA

studying whether the annoyance generated by such sound is systematically related to the physical parameters naturally varying under different operating conditions of the devices, such as sound pressure level (SPL) and signal-to-noise ratio, or to psychoacoustic metrics such as loudness and tonal content.

We decided that these issues were best addressed by taking the devices in question to the laboratory for precise acoustical measurements and binaural recordings in order to perform systematic listening tests on the sounds generated. The results of the listening tests will eventually constitute the yardstick by which any physical measurements will have to be weighted and evaluated. Therefore, these tests were designed with two very basic goals in mind: First, to determine, whether listeners are capable of consistently judging the sound quality of the inverters in the first place, and secondly to construct scales of auditory unpleasantness for a number of inverters under a wide range of operating conditions.

The first goal was adressed by collecting a complete set of paired comparisons on a limited number of sounds. It has been shown that if certain transitivity conditions (see [1]) are met by the judgments, a ratio scale of the subjective attribute in question may be constructed [2, 3]. Having determined the subjective scalability of the inverter sounds, the second goal was addressed by extending the scope of the listening tests to a larger number of devices and operating conditions using the more economic method of magnitude estimation [4, 5].

The purpose of the present study is thus threefold: (1) to collect reliable listening-test data based on two kinds of methodologies, (2) to simultaneously make acoustical measurements and compute psychoacoustical metrics (of loudness, tonal content, etc.) on the sounds under investigation, and finally (3) to relate the former to the latter, thus attempting to predict the auditory unpleasantness produced by the inverters using appropriate instrumental metrics derived from the signals.

2. Method

2.1. Participants

Fifty-three participants were recruited for the research. After audiometric screening in the range between 0.25 and 16 kHz, eight of them had to be excluded because their thresholds exceeded 20 dB hearing level (according to [6]) at one or more of the frequencies tested. Of the remaining 45 listeners (age 19–41 years; median age 22 years), 32 of whom were female, all participated in the paired-comparison experiment, and 42 in the magnitude-estimation experiment as well. A few listeners were staff members; the majority, however, were students who participated for a credit requirement in psychology.

2.2. Stimuli and apparatus

Five different solar inverters (in the following designated A–E) having maximal nominal power ranging from 2 kW (inverter E) to 15 kW (inverter A) and operated by an external unit were measured in different operating conditions inside a 5.5 m × 4.5 m × 4 m semi-anechoic chamber. To mimic their mounting to a wall they were attached to a metal rack covered by a quadratic 2.44 m × 2.44 m medium-density fiberboard plate.

Measurements for sound power calculations (details in [7]) were made using a hemispherical array of eight 1/4-inch microphones (Roga M-17) which – being mounted to a pivotable aluminum arch – could be rotated to five predetermined angles thus yielding 40 defined measurement positions placed at 1 m distance from the center of the inverter.

Recordings for playback in the listening tests were made in the same chamber using an artificial head (Head Acoustics HMS IV) positioned at 2 m distance from the inverters. The head was pointing at the source, but – to more effectively pick up sound radiation to the sides – was placed asymmetrically so that it formed a 45-degree angle with the panel on which the inverter was mounted.

Recordings were made while systematically varying the following operating conditions: (a) the power converted (25, 50, 75, or 100% of the maximum attainable), (b) the voltage fed to the inverter (typically a low and a high setting could be chosen), (c) the degree to which cooling ventilation was added (0, 25, 50, 75, 100%). In actual use, the power converted reflects turnover, i.e., solar input; the voltage depends on the number of solar panels fed to the inverter. As a result, 36 recordings were obtained from the five inverters. Four additional stimuli were generated by low-pass filtering the noise produced by one of the inverters (device A at 420 V) at 10 kHz in order to explore the role of the high-frequency components.

The artificial-head recordings were shortened to a uniform duration of 3 s including rise/fall times of 100 ms. They were D/A converted at 96 kHz and with 16 bit resolution by an external sound card (RME multiface II), passed through a headphone amplifier (Behringer Pro 8) and subsequently delivered to Beyerdynamic DT 990 headphones. Sound levels were calibrated using a 94-dB sinusoid the level of which was measured at the headphones using an artificial ear (Brüel & Kjær type 4153) connected to a sound level meter (Brüel & Kjær type 2250). The playback level was raised by 6 dB relative to the recording for less effortful listening to the sometimes very soft sounds, and to make the artifical-head recordings from 2 m distance roughly comparable to the parallel microphone recordings made at 1 m distance.

2.3. Procedure

Participants were run individually in a double-walled sound-attenuated (Industrial Acoustic Company) chamber. They were told that the experiment was about noise from solar-power equipment, and that they would have to judge potentially annoying sounds in two kinds of procedures. After obtaining informed consent, the participants first completed all paired comparisons of a selection of 16

inverter sounds, followed by magnitude estimates on all 40 sounds under investigation.

2.3.1. Paired comparisons

Sixteen sounds – none of which included noise from the cooling ventilation – were selected for paired comparisons: Inverter A both at 750 and 420 V, the latter in the low-pass filtered version as well, and inverter D, each of them presented at four power settings (25, 50, 75, and 100%). Inverter A was the one on which measurements in all operating conditions had been made, and inverter D was selected for comparing with a different brand. Each subject was presented with all possible $n(n-1)/2 = 120$ pairs of the 16 sounds presented in random order. Before starting the listening test, subjects were alerted to the difference between loudness and unpleasantness/annoyance by pointing out that there may be soft sounds that are very unpleasant.

A trial consisted of the presentation of a pair of sounds separated by a 0.5-s pause. The onset of each sound was accompanied by the appearance of a circle on the screen, to the left of the center and marked 'A' for the first, and to the right of the center and marked 'B' for the second observation interval. Subjects were instructed to click on the circle (A or B) corresponding to the sound that appeared 'more unpleasant' (German: 'unangenehmer') to them. They had a chance to repeat the presentation of a pair by clicking a button. Once they made their response, a 1-s inter-trial interval elapsed after which the next pair of sounds was presented. There were two rest breaks after 40 trials each. Completing all 120 paired comparisons took approximately 35 min.

2.3.2. Magnitude estimation

Following a short break, the participants were asked to judge the entire sample of sounds ($n = 40$, see section 2.2) using a magnitude estimation paradigm with a fixed standard (i.e., ratio estimation, [5]). On each trial they were presented with a pair of sounds: The standard, for which a scale value of '10' was agreed upon, followed – after an inter-stimulus interval of 0.5 s – by the comparison to be judged. Based on informal listening, the standard was chosen from the middle of the unpleasantness range (inverter B at 750 V and 50% power). The instructions explained how to assign numbers to reflect ratios of subjective magnitudes, and subjects entered their judgements into a field on the screen using the numeric keypad. Each participant judged each of the 40 sounds presented in random order once.

Subjects were run individually and took approximately one hour to complete the entire experiment (hearing test, paired comparisons, magnitude estimation).

3. Results and discussion

3.1. Paired comparisons

Before further processing, each individual data set was checked for inconsistencies in the paired comparisons known as circular triads: $a > b$, $b > c$ but $a < c$. Their number ranged from 0 to 65 out of a maximum possible $T_{\max} = 168$ circular triads [8], with a median value of $T = 14$. Kendall's coefficient of consistency $\zeta = 1 - T/T_{\max}$ was computed for each participant, and all those yielding a $\zeta < 0.80$, i.e., five subjects generating 38 to 65 circular triads, were excluded from further analysis.

The remaining 40 subjects' data were pooled and further analyzed using the 'eba' statistical package in the free software environment 'R' [9]. The aggregated data may be interpreted as probabilities for judging one sound to be more unpleasant than another, and can be checked for transitivity on the basis of the entire sample [10]. They exhibited no single violation of weak stochastic transitivity (WST), meaning that if both the probability of judging sound a to be more unpleasant than b was $p_{ab} \geq 0.5$, and the probability of finding b to be more unpleasant than c was $p_{bc} \geq 0.5$, then the probability of judging a to be more unpleasant than c was also $p_{ac} \geq 0.5$, implying that a consistent ordering of the stimuli is possible based on the aggregated paired comparisons. They further exhibited moderate stochastic transitivity (MST) in that, given the same two premises, the stronger conclusion $p_{ac} \geq \min(p_{ab}, p_{bc})$ held with just 4 violations in 560 tests. By contrast, strong stochastic transitivity (SST), which requires $p_{ac} \geq \max(p_{ab}, p_{bc})$, was violated in 81 of 560 tests. This pattern of transitivities suggests that the pooled data are quite consistent, but that the highly restrictive Bradley-Terry-Luce (BTL) model – requiring SST to hold – may not be adequate to describe them.

Consequently, a maximum-likelihood test [9] for the adequacy of the BTL model showed small but significant violations, $\chi^2(105) = 131.3$, $p = 0.042$, meaning that it fails to describe all paired comparisons under study. When, however, the inverter producing the greatest annoyance (device A, run at 700 V and 100% power) was removed from the analysis, the deviations from the model were no longer significant, $\chi^2(91) = 106.5$, $p = 0.127$. Note that the excluded stimulus was judged to be more unpleasant than any other stimulus, in many cases by all participants, thus making it hard to estimate its perceptual distance from the others. Thus the failure to include this stimulus in the model is seen as a technical rather than a substantial issue.

Based on the BTL model, scale values for the 15 remaining sounds under study were estimated [9]. They are depicted in Figure 2. It is evident that for all inverters scaled auditory unpleasantness increases monotonically with the power converted. Inverter D is distinguished from the others in that its unpleasantness varies less as a function of power, while starting out with higher scale values at the lowest power setting already.

The conclusion to be drawn from this part of the investigation is that – with one exception – paired comparisons of the 16 inverter sounds studied exhibit enough structure to be ratio-scalable. Given the fact that listeners appear to be capable of assessing unpleasantness on an internal ratio scale, a larger sample of sounds was studied using direct magnitude estimates of auditory unpleasantness.

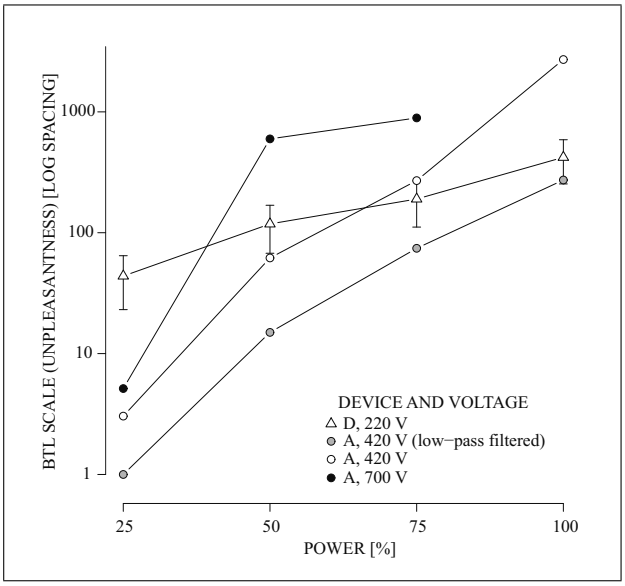

Figure 2. BTL-scale values of auditory unpleasantness derived from the paired comparisons. Error bars represent 95-% confidence intervals of the estimates for device D (triangles). Open circles represent scale values for device A at 420 V; grey, filled circles denote the same sounds when low-pass filtered at 10 kHz. Black, filled circles refer to device A run at 700 V, with the highest power level (100%) not representable by a BTL model.

3.2. Magnitude estimation

In order to preserve the ratios and to minimize the effect of outliers, the magnitude estimates[1] provided by the 42 normal-hearing participants were geometrically averaged. The resulting unpleasantness values are depicted in Figure 3 separately for each inverter (A–E) as a function of the power at which the equipment was run (left part of the figure) and as a function of the extra cooling ventilation added at full power (right part of the figure). It is evident that the standard errors of the geometric means (computed on the basis of the logarithms of the raw data and hence slightly asymmetric in the plot) are relatively small.

Furthermore, the inverters vastly differ in the overall annoyance produced, with device C (bottom curve; open and filled diamonds in Figure 3) producing the lowest unpleasantness ratings. Due to the unbalanced experimental design, the initial statistical analyses of overall effects were conducted by computing separate, one-factor, repeated-measures analyses of variance [2] (ANOVAs). The first such analysis confirmed that the different makes of inverters produce highly significant differences in the auditory unpleasantness they evoke, $F(4, 164) = 107.67$; $p < 0.001$ ($M_C = 6.6$; $M_D = 12.5$; $M_B = 20.1$; $M_A = 24.2$; $M_E = 35.7$). More importantly, for all inverters (except C), perceived unpleasantness appears to monotonically grow with the power converted (as seen in the left part of Figure 3), resulting in a statistically significant main effect of power, $F(3, 123) = 161.72$; $p < 0.001$ ($M_{25\%} = 5.2$; $M_{50\%} = 10.0$; $M_{75\%} = 15.8$; $M_{100\%} = 31.3$). Adding ventilation (right part of Figure 3) further increases unpleasantness, nearly doubling the scores at the maximal (100%) setting. This effect is highly significant as well, though not in all instances monotonic with the amount of ventilation added, $F(4, 164) = 67.98$; $p < 0.001$ ($M_{\text{off}} = 9.1$; $M_{25\%} = 22.3$; $M_{50\%} = 29.8$; $M_{75\%} = 37.1$; $M_{100\%} = 29.3$)[3]. Finally, the recordings made with the higher voltage (for a given device: filled vs. open symbols) tend to result in higher unpleasantness estimates, resulting in an overall significant main effect of voltage, $F(7, 287) = 121.06$; $p < 0.001$ ($M_{200\,\text{V}} = 2.8$; $M_{220\,\text{V}} = 12.5$; $M_{250\,\text{V}} = 16.9$; $M_{420\,\text{V}} = 21.0$; $M_{470\,\text{V}} = 22.5$; $M_{650\,\text{V}} = 23.8$; $M_{700\,\text{V}} = 30.2$; $M_{750\,\text{V}} = 22.9$).

Note that the low-pass filtered signals (grey circles connected by dotted lines in the left panel of Figure 3) show systematically lower unpleasantness scores than their unfiltered counterparts (open circles connected by dashed lines in the same panel), with the difference growing as power is increased. An ANOVA performed on this subset of the data shows statistically significant main effects of both power, $F(3, 123) = 186.69$; $p < 0.001$, and filtering, $F(1, 41) = 75.46$; $p < 0.001$, as well as a significant (filtering × power) interaction, $F(3, 123) = 13.15$, $p < 0.001$. That implies that the unpleasantness ratings are significantly affected by the high-frequency tonal components emerging, particularly when the inverters operate near maximum power.

It is further remarkable that the shape of the curves and the relative ordering of conditions (notably for the low-voltage, high-voltage, and low-pass-filtered recordings of inverter A) is quite similar when comparing the outcome of the paired-comparison (Figure 2) and the magnitude-estimation experiment (Figure 3).

3.3. Psychoacoustic metrics

In order to evaluate whether standard instrumental metrics are suited to make predictions regarding perceived unpleasantness, such metrics [11] were computed from the binaural recordings using a sound-quality analysis software (Brüel & Kjær PULSE, Version 14). The software inversely filters the binaural signals with the artificial head's HRTFs to obtain the corresponding unobstructed free-field exposure and eventually provides 'monaural' metrics for the left and right ear. The values of these metrics were very close for the present recordings, and thus – with the exception of binaural loudness – averaged to result in a mean sound-quality score. In the latter case, an algorithm more heavily weighting the ear receiving the higher sound pressure level [12] was used. Different types of loudness metrics were computed (stationary, statistical, and binaural) as

[1] Three subjects had entered (mathematically impossible) values of zero in the ratio-estimation task once. These were replaced by values of 0.1.

[2] Due to the skewed distribution of the raw data, these analyses were performed on the logarithms of the ratio estimates.

[3] Surprisingly, device B at 250 V and 100% ventilation (the open square in the bottom right corner of Figure 3) is rated considerably less annoying than the same device with ventilation shut off. This anomaly is presently not understood, but is accounted for by the instrumental measurements (section 3.3), which indicate the same.

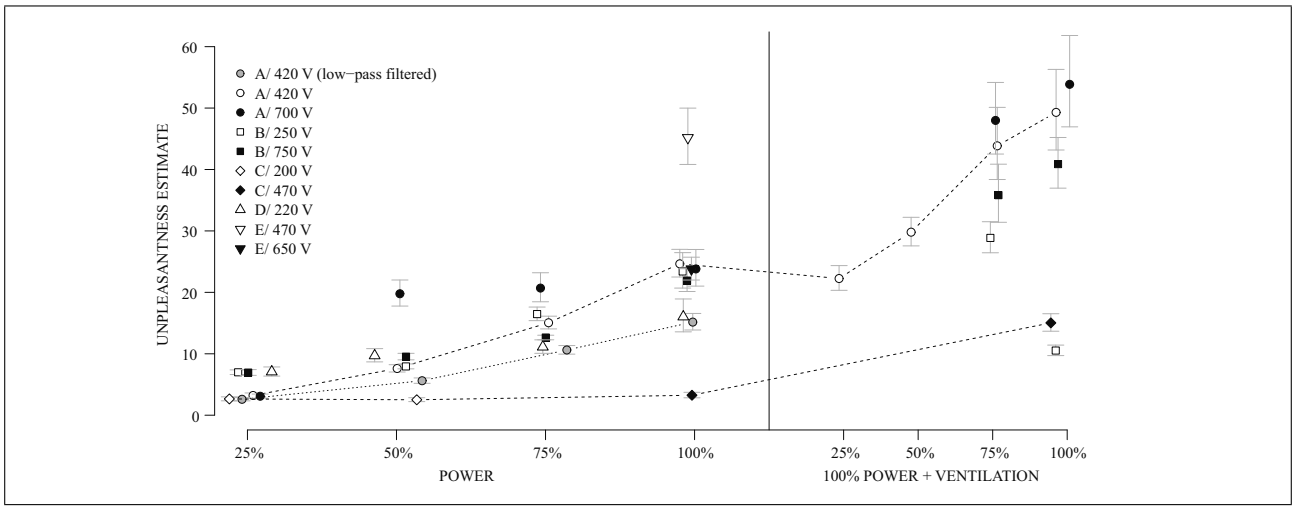

Figure 3. Perceived annoyance of the 40 inverter sounds studied as a function of the power converted (left) and the extra ventilation added at full power (right). Geometric means of the magnitude estimates made by 42 participants are depicted plus/minus one standard error of the mean. Measurements for selected devices (A and C) are connected by lines. For a given device, open symbols refer to recordings made with the lower, filled symbols to those made with the higher voltage setting. Circles filled in grey represent judgments of the low-pass filtered recordings. For better readability, the data points are slightly jittered around their proper *x*-axis values.

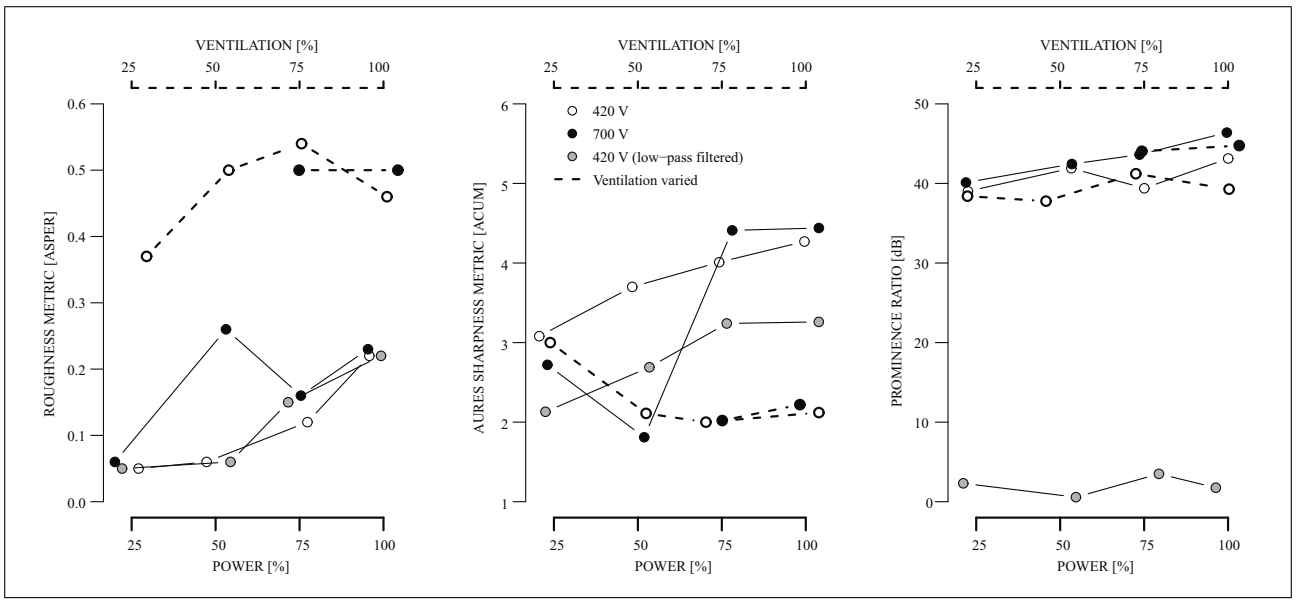

Figure 4. Instrumental measures of roughness (left), sharpness (center), and tonalness (right) computed for inverter A at various operating conditions, as indicated in the legend. Sounds including cooling ventilation refer to the upper abscissa and their data points are connected by dashed lines. The plotting symbols are slightly jittered around their *x*-axis values.

well as psychoacoustical sharpness, roughness, fluctuation strength, and the prominence ratio [13, 14] as a measure of tonal content.

The outcome is illustrated in Figure 4 for three sound-quality metrics (roughness, sharpness, and tonalness), using the recordings of inverter A for which the largest number of measurements was made. The figure uses a different representation than Figure 3 in that the data for the ventilation conditions (all of which were collected with the inverter running at 100% power) are plotted along a second abscissa at the top of the figure and connected by dashed lines. Psychoacoustical roughness (left panel of Figure 4), reflecting rapid frequency or amplitude modulations [11], shows a similar increase with the power converted as did the subjective annoyance estimates (Figure 3), and is further increased by ventilation noise (dashed lines in Figure 4). Sharpness, computed based on Aures' method [15], also tends to increase with operating power (center panel of Figure 4), but here, ventilation has the opposite effect: It reduces the sharpness produced by the high-frequency tones by shifting the spectral balance downward in frequency (dashed lines), as does low-pass filtering (symbols filled in grey). The prominence ratio, reflecting to what extent one or several tonal components exceed the level of the neighboring critical bands [13], is very high in all actual recordings (right panel of Figure 4), with a slight

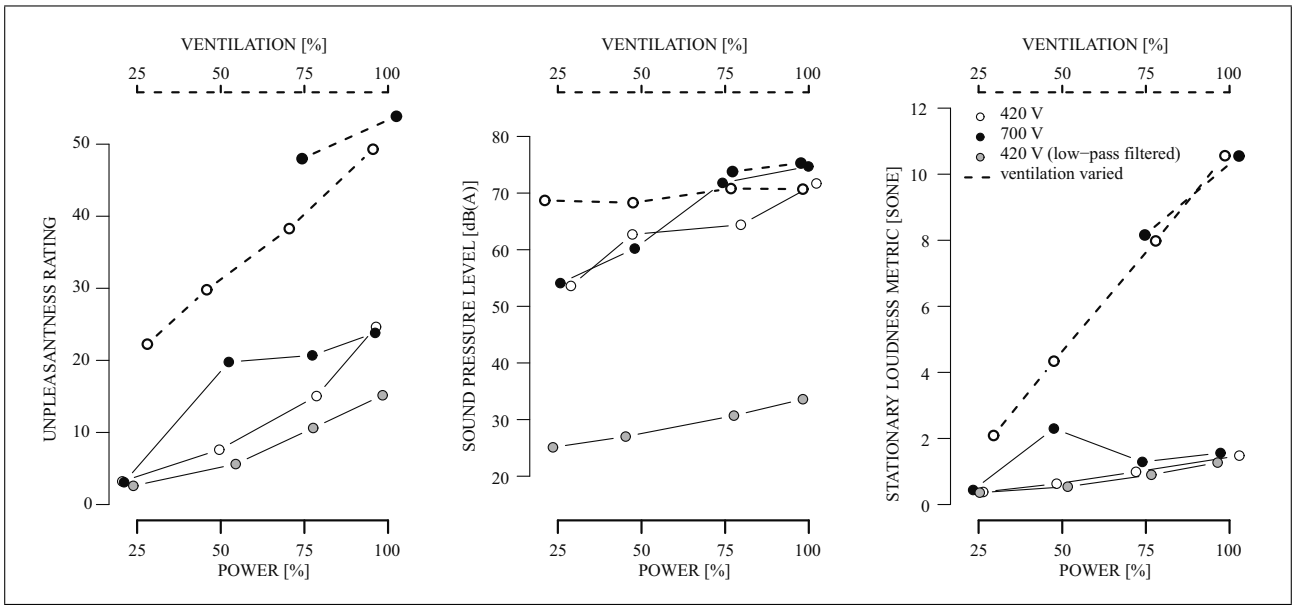

Figure 5. Unpleasantness ratings (left), A-weighted sound pressure levels (center), and instrumentally measured stationary loudness (right) of inverter A in different operating conditions. The lower abscissa represents the power at which the inverter was run with ventilation turned off and refers to all data connected by solid lines; the upper abscissa refers to the data points connected by dashed lines and shows the amount of additional ventilation while the equipment ran at 100% power.

tendency to increase with power. As expected, it is almost reduced to zero for the low-pass filtered sounds, and the presence of ventilation does not seem to affect tonal prominence.

Thus, the psychoacoustic metrics analyzed so far all appear to show some covariance with auditory unpleasantness. Which ones predict the latter best will be analyzed in the next section.

The most likely psychoacoustic metric to have an influence on the unpleasantness of the inverter sounds, however, is loudness. In Figure 5 stationary ('Zwicker', i.e. DIN 45631) loudness (right panel) is contrasted with the mean unpleasantness ratings (left panel) assigned by the listeners and with A-weighted sound pressure level (center panel). It is evident that instrumentally measured loudness and subjectively scaled unpleasantness produce qualitatively quite similar curves, with the notable exception that taking out the high-frequency tonal component at 18 kHz (by low-pass filtering the sounds, see the bottom curves in Figure 5) hardly affects loudness, while having a small but significant effect on unpleasantness. That may be due to the critical-band filters used for loudness computation having an upper limit at 15.5 kHz [11] and the high-frequency components being present in some of the signals, thus affecting loudness by the tails of their excitation patterns only.

It is interesting to contrast the loudness metric with physical measurements of A-weighted sound pressure level (see the center panel of Figure 5): While overall sound pressure levels also increase with power, (a) the low-pass filtered signals drop by some 30 dB with respect to the unfiltered originals, and (b) the added ventilation has practically no effect on this measure. Both of these effects are due to the high-frequency tonal components (which

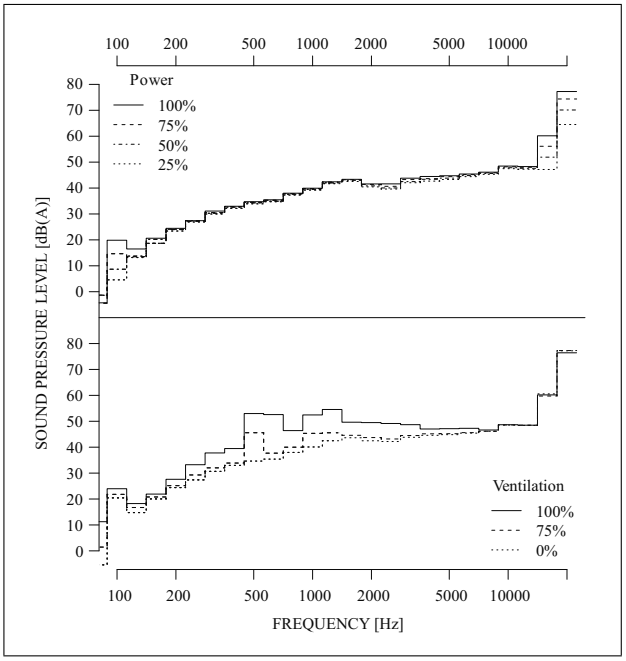

Figure 6. Third-octave band spectra of inverter B as a function of the power converted (top) and of the added ventilation when running at the highest power level (bottom).

are elevated by some 30–40 dB with respect to the remainder of the spectrum, see Figure 1) dominating the overall sound pressure levels.

The discrepancy between loudness and sound-pressure level is further elucidated by inspecting the spectra of the inverters as a function of frequency and comparing the effects of turning up power versus increasing ventilation (Figure 6). While the former almost exclusively influences the high-frequency peaks (top graph), the latter affects the

Table I. Correlations between various signal-based measures and auditory unpleasantness. $^{**}p < 0.01$; $^{***}p < 0.001$.

Sound pressure level	SPL [dB(A)]	SPL [dB(lin)]	SPL [dB(mic)]
	0.675***	0.664***	0.615***
Loudness metrics	N_{bin} [sone]	N [sone]	N_5 [sone]
	0.904***	0.897***	0.898***
Sound quality metrics	S_{Zwicker} [acum]	S_{Aures} [acum]	FS [vacil]
	−0.242	−0.123	0.835***
	R [asper]	T/N [dB]	Prominence [dB]
	0.868***	0.219	0.418**

mid-frequency region of the spectra (bottom graph) and thereby has negligible effects on the overall level. It thus appears that, while level-based measures (like A-weighted sound pressure level) are dominated by the elevated high-frequency tonal components, subjective as well as sound-quality indices are much more sensitive to the effect of the ventilation noise operating in the mid-frequency range (400–4000 Hz).

3.4. Relation between metrics and unpleasantness

Table I summarizes the correlations between perceived unpleasantness on the one hand and a number of psychoacoustical metrics (including a few measures of sound pressure level) on the other. It is evident – by squaring the correlation coefficients obtained – that just on the basis of a single loudness measure, over 80% ($r^2 = 0.904^2 = 0.82$) of the variance in unpleasantness ratings is accounted for, whereas any measure of sound pressure level accounts for approximately half as much ($r^2 = 0.675^2 = 0.46$).

The correlations of other sound-quality metrics with unpleasantness vary considerably; therefore, an attempt was made to assess their combined predictive potential by using a multiple-regression approach. The statistical rationale chosen (forward selection, see [16]) first identifies that predictor that accounts for most of the variance and then successively enters new predictors into the regression equation, provided they improve the prediction significantly. Initially, three different sound pressure level measurements (linear and A-weighted SPLs based on the artificial-head recordings as well as SPLs derived from the microphone-array measurements [SPL(mic)]; see the top entries in Table I) and ten different psychoacoustic metrics (of loudness, sharpness, roughness, fluctuation strength, and tonal content; remaining entries in Table I supplemented by loudness as determined by averaging metrics obtained from the 40 microphone positions in the array) were considered.

The statistical procedure eliminated all sound-pressure level based predictors, and suggests a purely psychoacoustical model. With the requirement that only variables contributing more than 5 percent of predictive power be included, the following regression model resulted,

$$U = -7.73 + 3.36 \cdot N_{\text{bin}} + 36.2 \cdot R + 3.08 \cdot S_{\text{A}}, \qquad (1)$$

where U is the auditory unpleasantness to be predicted, N_{bin} is binaural loudness as defined in [12], R is psychoacoustical roughness, and S_{A} is sharpness (the spectral centroid) based on Aures' algorithm [11, 15].

This regression model accounts for approximately 94% of the variance (corrected R^2) with loudness contributing approximately 81.3% and roughness and sharpness contributing approximately 6% each (5.8 and 6.4%, respectively).

4. General Discussion

The listening tests reported in the present study show (1) that listeners can consistently judge the psychoacoustical unpleasantness of the sounds radiated in the vicinity of photovoltaic equipment, (2) that this auditory unpleasantness can be quantified both using direct and indirect psychophysical scaling methodology, and (3) that the outcome is meaningfully related to instrumental metrics of sound quality derived from the acoustical signals. In the following we will elaborate these points and discuss their implications for the assessment of sound quality.

4.1. Validity and scale type of unpleasantness scores

A sampling of the present stimulus set – manageable in terms of performing all possible paired comparisons – turned out to be representable by the highly restrictive BTL model, provided the sound yielding the highest unpleasantness scores was excluded. That is very likely due to the indeterminacy of its scale value (being judged more unpleasant with relative frequencies of 100% with respect to most other sounds), rather than its sound quality. Note that attempting other representations in terms of the less restrictive preference tree model [17], i.e., defining subgroups of stimuli (e.g., by make or type of inverter) potentially governed by other sound features, did not yield a better model fit.

Other psychoacoustical studies using paired comparisons have yielded similar results: A study of a heterogeneous collection of environmental sounds found a BTL model to adequately describe the data [18] as did a study determining preferences for various spatial audio reproduction formats [19]. By contrast, another study from our laboratory [20] required modeling by a preference tree, distinguishing subgroups of loud vs. soft sounds, judged by different criteria.

Both a conceptual and a practical advantage of the BTL model with respect to a preference tree representation is

that the BTL model implies unidimensionality, i.e., all stimuli being compared may be assumed to be judged by the same criterion or set of criteria. Only given such unidimensionality is it justified to propose a single regression equation predicting overall unpleasantness as in equation (1), otherwise, identifiable subgroups of stimuli would have to be characterized by different sets of psychoacoustical criteria or regression weights (as in [20]). Note, however, that unidimensionality was only determined paradigmatically for the 16 (15) stimuli included in the paired-comparison experiment, not for all 40 stimuli later subjected to direct scaling.

If the v-scale derived from the BTL model is by definition [2] a ratio scale of unpleasantness, the fact that it yields perceptual ratios approaching 1:2000 is somewhat puzzling. Indeed, most investigators (e.g., [20, 21]) have observed that the logarithm of the 'raw' BTL scale values more closely corresponds to what people would produce as perceptual ratios in a direct estimation task. In the present investigation, the BTL scale values only moderately correlated with the subsequent magnitude estimates of the same 15 sounds at $r = 0.775$, while their logarithms showed a much higher correlation with the direct estimates: $r = 0.945$ (compare Figures 2 and 3).

4.2. Predictive value of physical and psychoacoustic metrics

The present study of a relatively homogeneous set of machinery noises showed a direct-scaling measure of auditory unpleasantness obtained using Stevens' ratio estimation with a fixed standard [4, 5] to produce results of high inter-individual consistency that exhibit systematic dependencies on the type and operating conditions of the devices assessed. Since the direct scaling technique allowed a much larger number of stimuli to be studied ($n = 40$ sounds) than the BTL technique largely included for validational purposes ($n = 16$ sounds), it was possible to relate the scaling results to objectively defined psychoacoustic metrics on the basis of a more sizeable sample of inverter sounds. That led to a linear regression model, according to which three instrumentally measured psychoacoustical indices (loudness, roughness, and sharpness, see equation (1)) serve to predict the perceived unpleasantness of the sounds with great accuracy.

The most noteworthy result in our view is that the best prediction turned out to be purely psychoacoustical, and that physical measures such as sound pressure level or A-weighted level did not need to be included, and generally failed to make a good prediction. That is consistent with the general conclusion drawn in the literature on unpleasantness/annoyance [11, 22, 23, 24] that psychoacoustically motivated metrics (such as instrumentally measured loudness) fare better in predicting human auditory responses than level-based measures (such as A-weighted decibels). Note, however, that most legislation and guidelines regarding environmental noise or product sound are based on level-based measures (L_{eq} or L_A), which go a long way if sound pressure level highly correlates with loudness (and annoyance) which it very often does. That was not the case for the present stimulus material, since increasing the operating power of the devices raised the high-frequency peaks by some 20 dB (see Figure 6, top panel) while unpleasantness and loudness (Figure 5) were affected to a much smaller extent. On the other hand, adding more ventilation, and thereby mid-frequency noise (see Figure 6, bottom panel), strongly affected loudness and unpleasantness, while the overall level hardly changed (Figure 5), being dominated by the high-frequency peaks. The scaling results show that our listeners perceived the high-frequency components but that their effects were not as strong as the decibel difference would suggest. It appears that even when an A-weighting, attenuating the influence of high frequencies, is applied, level-based measures of annoyance overestimate the effect of the prominent tonal components close to the limit of the human frequency range.

The present set of recordings from actual photovoltaic inverters was purposely not loudness-equalized to study the natural sound levels produced by these devices. As is common under such circumstances, loudness emerged as by far the most influential factor accounting for perceived unpleasantness/annoyance. Often, only when loudness is controlled for, do other psychoacoustic metrics become significant [25]. In the present case, however, the optimal prediction did include other psychoacoustic metrics as well, namely, roughness and sharpness, each making minor contributions to overall unpleasantness. That kind of model very much agrees with proposals for a generic, domain-independent formula predicting annoyance such as Fastl and Zwicker's 'psychoacoustic annoyance' [11].

4.3. Conclusions

The noise produced by photovoltaic inverters varies considerably both in level and quality. Nevertheless, paired comparisons of inverter sounds with respect to their annoyance exhibited remarkable consistency, permitting construction of a BTL scale for most of the stimuli, and suggesting that annoyance is ratio-scalable. Measurements made on a larger set of inverter sounds using direct magnitude estimation further demonstrated meaningful relationships with the operating conditions (e.g., power converted, amount of cooling ventilation) under which the equipment was recorded. Finally, instrumental measures of sound-quality parameters computed from the signals directly were shown to account for over 90% of the variance of subjective annoyance estimates, suggesting a reasonable prediction based on properties of the acoustic stimuli alone is feasable.

The main practical lesson, however, is that A-weighted sound pressure level would constitute an unsuitable predictor of the annoyance produced by photovoltaic inverters. Psychoacoustically motivated instrumental measures fare much better in this respect with stationary loudness alone making a fairly good prediction, and a weighted combination of several metrics performing even better.

Acknowledgements

We would like to thank Maria Roos at Fraunhofer Institute for Wind Engergy and Energy Systems IWES, Kassel, Germany, and Steffen Seeber from voltwerk electronics (now with Bosch Power Tec, Bad Vilbel, Germany) for taking the initiative to start this research, for providing the photovoltaic equipment, and for engaging in continuing discussions on the psychoacoustical issues involved. Portions of the data were presented at the DAGA 2011 (Düsseldorf, Germany) and Forum Acusticum 2012 (Aalborg, Denmark) meetings.

References

[1] R. D. Luce, P. Suppes: Preference, utility, and subjective probability. – In: Handbook of mathematical psychology, Vol. III. R. Luce, R. Bush, S. Galanter (eds.). Wiley, New York, 1965, 249–410.

[2] R. D. Luce: Individual choice behavior. Wiley, New York, 1959.

[3] K. Zimmer, W. Ellermeier: Deriving ratio-scale measures of sound quality from preference judgments. Noise Control Engineering Journal **51** (2003) 210–215.

[4] S. S. Stevens: Psychophysics: Introduction to its perceptual, neural, and social prospects. Transaction, Piscataway, NJ, 1975.

[5] G. A. Gescheider: Psychophysics. The fundamentals. 3rd ed. Lawrence Erlbaum, Mahwah, NJ, 1997.

[6] ISO: Acoustics - reference zero for the calibration of audiometric equipment - part 2: Reference equivalent threshold sound pressure levels for pure tones and supraaural earphones (ISO 389-1). International Organization for Standardization, 1998.

[7] A. Meister, K. Moritz, A. Skowronek, L. Kurtze, W. Ellermeier, M. Roos, S. Seeber: Akustische Messungen von Photovoltaik-Wechselrichtern. DAGA 2011, Düsseldorf, 2011, Deutsche Gesellschaft für Akustik, 193–194.

[8] M. G. Kendall, B. B. Smith: On the method of paired comparisons. Biometrika **31** (1940) 324–345.

[9] F. Wickelmaier, C. Schmid: A Matlab function to estimate choice model parameters from paired-comparison data. Behavior Research Methods, Instruments, and Computers **36** (2004) 29–40.

[10] P. Suppes, D. H. Krantz, R. D. Luce, A. Tversky: Foundations of measurement. Vol. II. Academic Press, San Diego, 1989.

[11] H. Fastl, E. Zwicker: Psychoacoustics - Facts and Models. 3rd ed. Springer, Berlin, 2007.

[12] V. P. Sivonen, W. Ellermeier: Binaural loudness for artificial-head measurements in directional sound fields. Journal of the Audio Engineering Society **56** (2008) 452–461.

[13] M. Nobile, G. Bienvenue: Procedure for determining the prominence ratio of discrete tones in noise emissions. Noise Con 91, Tarrytown, NY, 1991, Institute of Noise Control Engineering of the USA, 475–482.

[14] H. Hansen, J. L. Verhey, R. Weber: The magnitude of tonal content – a review. Acta Acustica united with Acustica **97** (2011) 355–363.

[15] W. Aures: Berechnungsverfahren für den sensorischen Wohlklang beliebiger Schallsignale. Acustica **59** (1985) 130–141.

[16] I. H. Bernstein, C. P. Garbin, G. K. Teng: Applied multivariate analysis. Springer, New York, 1988.

[17] A. Tversky, S. Sattath: Preference trees. Psychological Review **86** (1979) 542–573.

[18] W. Ellermeier, M. Mader, P. Daniel: Scaling the unpleasantness of sounds according to the BTL model: Ratio-scale representation and psychoacoustical analysis. Acta Acustica united with Acustica **90** (2004) 101–107.

[19] S. Choisel, F. Wickelmaier: Evaluation of multichannel reproduced sound: Scaling auditory attributes underlying listener preference. Journal of the Acoustical Society of America **121** (2007) 388–400.

[20] K. Zimmer, W. Ellermeier, C. Schmid: Scaling the unpleasantness of sounds according to the BTL model: Ratio-scale representation and psychoacoustical analysis. Acta Acustica united with Acustica **90** (2004) 1019–1028.

[21] M. Marzinzik, B. Kollmeier: Predicting the subjective quality of noise reduction algorithms for hearing aids. Acta Acustica united with Acustica **89** (2003) 521–529.

[22] H. Fastl: Psychoacoustics and sound quality. – In: Communication Acoustics. J. Blauert (ed.). Springer, Berlin, 2005, 139–162.

[23] C. Marquis-Favre, E. Premat, D. Aubrée: Noise and its effects – A review on qualitative aspects of sounds. Part II: Noise and annoyance. Acta Acustica united with Acustica **91** (2005) 626–642.

[24] T. Kaczmarek, A. Preis: Annoyance of time-varying road traffic noise. Archives of Acoustics **35** (2010) 383–393.

[25] S. More: Aircraft noise characteristics and metrics. PhD dissertation. Purdue University, 2010.