

COVER PAGE

Document downloaded by @DAEL

Fri Nov 27 21:01:23 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

Spectro-Temporal Characteristics Affecting the Loudness of Technical Sounds: Data and Model Predictions, Jan Rennies, Moritz Wächtler, Jan Hots and Jesko Verhey, Acta Acustica **vol. 101** (Number 6), 2015, pp. 1145-1156

DOI
https://doi.org/10.3813/AAA.918907

ACTA ACUSTICA UNITED WITH ACUSTICA
Vol. 101 (2015) 1145 – 1156
DOI 10.3813/AAA.918907

Spectro-Temporal Characteristics Affecting the Loudness of Technical Sounds: Data and Model Predictions

Jan Rennies[1)], Moritz Wächtler[1)], Jan Hots[2)], Jesko Verhey[2)]
[1)] Cluster of Excellence Hearing4all Project Group Hearing, Speech and Audio Technology, Fraunhofer Institute for Digital Media Technology IDMT, D-26129 Oldenburg, Germany. jan.rennies@idmt.fraunhofer.de
[2)] Department of Experimental Audiology Otto von Guericke University Magdeburg Leipziger Str. 44, 39120 Magdeburg, Germany

Summary
Technical sounds are often non-stationary, i.e., their levels vary over time. Thus, temporal as well as spectral characteristics influence the loudness of these sounds. The present study investigated the loudness of technical sounds with a special focus on sounds with pronounced temporal fluctuations. Levels at equal loudness between these sounds and a reference sound were measured for normal-hearing listeners using an adaptive matching procedure. A jet noise was used as the reference sound. In addition to real technical sounds, loudness was measured for stimuli which preserved only the temporal or only the spectral structure of an original technical sound. The results were compared with predictions of published dynamic loudness models and predictions based on the German and American standards for loudness. The accuracy of the models depended strongly on the type of sound. In general, predictions were least accurate for sounds with large temporal fluctuations. The comparison of the loudnesses of the original sounds with those, where only the spectrum or the temporal structures were preserved, revealed that spectral characteristics often dominate loudness. Although none of the tested models could predict the loudness of all employed time-varying, technical signals, the longer time-constants used in one of the models appear to be beneficial.

PACS no. 43.66.Cb, 43.66.Ba, 43.50.Rq, 43.66.Mk, 43.50.Jh

1. Introduction

The perceived intensity of a sound, i.e., its loudness, has a strong influence on the assessment of environmental noise. For example, the loudness substantially contributes to the annoyance of a sound [1]. Thus, loudness is one of the fundamental auditory sensations and methods for calculating loudness have been described in national and international standards [2, 3, 4]. Although loudness is correlated with the physical property sound intensity, it also depends on several other characteristics of the sound, such as the temporal envelope and the spectrum of the sound. Not all of these aspects are currently accounted for in the standards. For example, ANSI S3.4 and ISO 532 use long-term spectra as input for the calculation of the loudness. The same is true for DIN45631. Thus, these standards focus on the influence of spectral properties. These standards take into account the fact that loudness depends on frequency as reflected in the equal-loudness level contours (e.g., [5, 6]). They also include stages that account for spectral loudness summation. Spectral loudness summation describes the effect that broadband sounds are in general louder than equal-level narrowband sounds. The effect is accounted for by assuming that the auditory system analyzes the incoming sound by a bank of overlapping bandpass filters. The output of each filter is compressed and then summed across frequency bands. The exact parameters of the stages differ between the standards. For example, DIN45631 uses the critical bandwidth values given in [7], whereas ANSI S3.4 uses the equivalent-rectangular-bandwidth (ERB_N) as proposed in [8]. Both estimates of auditory filter width were derived from psychoacoustical masking data. The estimates differ because different masking paradigms were used to determine the frequency selectivity of the auditory system. Recent data measuring frequency selectivity using notched-noise paradigms [9, 10] indicate that auditory filters are narrower than the bandwidths indicated by Zwicker [7] and, hence, closer to ERB_N. However, which of the two estimates is more appropriate for the prediction of loudness of broadband sounds is still a matter of discussion [11]. Environmental sounds are commonly broadband, i.e., a realistic prediction of spectral aspects of loudness is important when assessing their loudness.

Environmental sounds are not always stationary, i.e., their levels and spectral shapes may vary over time. Sev-

Received 13 March 2015,
accepted 10 July 2015.

© S. Hirzel Verlag · EAA

Table I. Overview of the signals used in this study. Check marks indicate that the signal was used in the respective experiment.

No.	Name	Exp. 1	Exp. 2	Exp. 3	Reference
1	jet lin (reference)	✓	✓	✓	Gee et al. (2007) [29]
2	hammer	✓	✓	✓	freesound: cognito perceptu, 17012 [30]
3	helicopter	✓	✓	✓	freesound: klankbeeld, 116692 [30]
4	machine gun	✓	✓	✓	freesound: Matt_G, 30749 [30]
5	ratchet	✓	✓	✓	freesound: atomota, 26869 [30]
6	snare drum	✓	✓	✓	own recording
7	impact wrench	✓		✓	Brennecke and Remmers (1983) [27]
8	planing machine	✓		✓	Brennecke and Remmers (1983) [27]
9	screw wrench	✓		✓	freesound: digifishmusic, 54264 [30]
10	jet non lin	✓			Gee et al. (2007) [29]
11	anvil		✓		freesound: l0calh05t, 19020 [30]
12	diesel engine		✓		freesound: hanstimm, 7803 [30]
13	gun powder		✓		freesound: Jon285, 34708 [30]
14	locomotive		✓		freesound: Sonic, 2881 [30]
15	ship horn		✓		Johannsen and Prante (2001) [28]

eral experiments with artificial sounds showed that the temporal properties of a sound affect its loudness. In general, short-duration sounds require a higher intensity to match the loudness of equally-loud long-duration sounds. This effect, commonly known as temporal integration of loudness, is often described using a low-pass filtering of instantaneous loudness values with a time constant of about 100 to 200 ms, followed by a stage calculating the maximum or a percentile as an estimate of the overall loudness of the sound [12]. Some researchers have argued that more than one time constant is required to predict the data [13, 14, 15] and models such as the one described in DIN 45631/A1 [16] or as proposed by Glasberg and Moore [17] include a temporal integration stage with several time constants.

Loudness also depends on the rapidity of level fluctuations. The loudness of a sinusoidally amplitude modulated sound is usually higher than that of a stationary sound with the same intensity. The level difference at equal loudness of these sounds is largest for modulation frequencies of a few Hertz (e.g., [18, 19, 20]). Compared to spectral loudness summation and temporal integration of loudness the effect is rather small (only a few decibels). Some models such as the dynamic loudness model (DLM, [21]) use the same model structure to predict temporal integration of loudness and the loudness of amplitude modulated sounds. In contrast, the model for time-varying loudness (TVL, [17]) uses a long-term loudness to predict the loudness of modulated sounds and the short-term loudness to predict temporal integration of loudness. The time constants for the long-term loudness are longer than those for the short-term loudness and the overall loudness is calculated as the mean instead of the maximum of the loudness-time function.

Some studies indicate that spectral and temporal integration of loudness are not independent processes. Spectral loudness summation depends on duration [22, 23]. This effect is also observed for sequences of short noise bursts [24] and influences spectral integration for temporally non-overlapping frequency components [25]. Rennies and colleagues [26] proposed an extended version of the DLM with a bandwidth-dependent onset enhancement that partly accounted for this spectro-temporal aspect of loudness. Whereas some temporal aspects of loudness are included in one standard (DIN45631), more complex spectro- temporal effects have not yet been considered in loudness standards. It is unclear if these effects play a role in the loudness of environmental sounds.

The present study investigates the loudness of technical sounds with a special focus on sounds with large temporal fluctuations. In particular, the aim was to understand the influence of the spectral and temporal properties of sounds on loudness. To this end, original technical sounds as well as artificial sounds which only preserved the spectral aspects or the temporal aspects of the original sounds were used. The measured loudness was compared to predictions of various standards and loudness models.

2. Methods

2.1. Stimuli

Fifteen original technical sounds were used. The waveforms are shown in Figure 1; insets in each panel represent long-term one-third-octave spectra.

Table I lists all signals and their sources [27, 28, 29, 30]. Only a subset of the stimuli was used in the three experiments of this study. Check marks indicate the experiments in which they were used. The nine technical signals used in Experiment 1 ranged from rather stationary (e.g., *jet linear*) to highly impulsive signals (e.g., *machine gun*). Some signals were taken from previous studies (see last column); the remaining signals were taken from a free online sound data base, and one signal was a recording made by one of the authors. The durations of the signals varied between 1.60 and 2.62 s. The signal selection for Experiment 1 was motivated by predictions of the tested models and aimed at including signals with considerably different

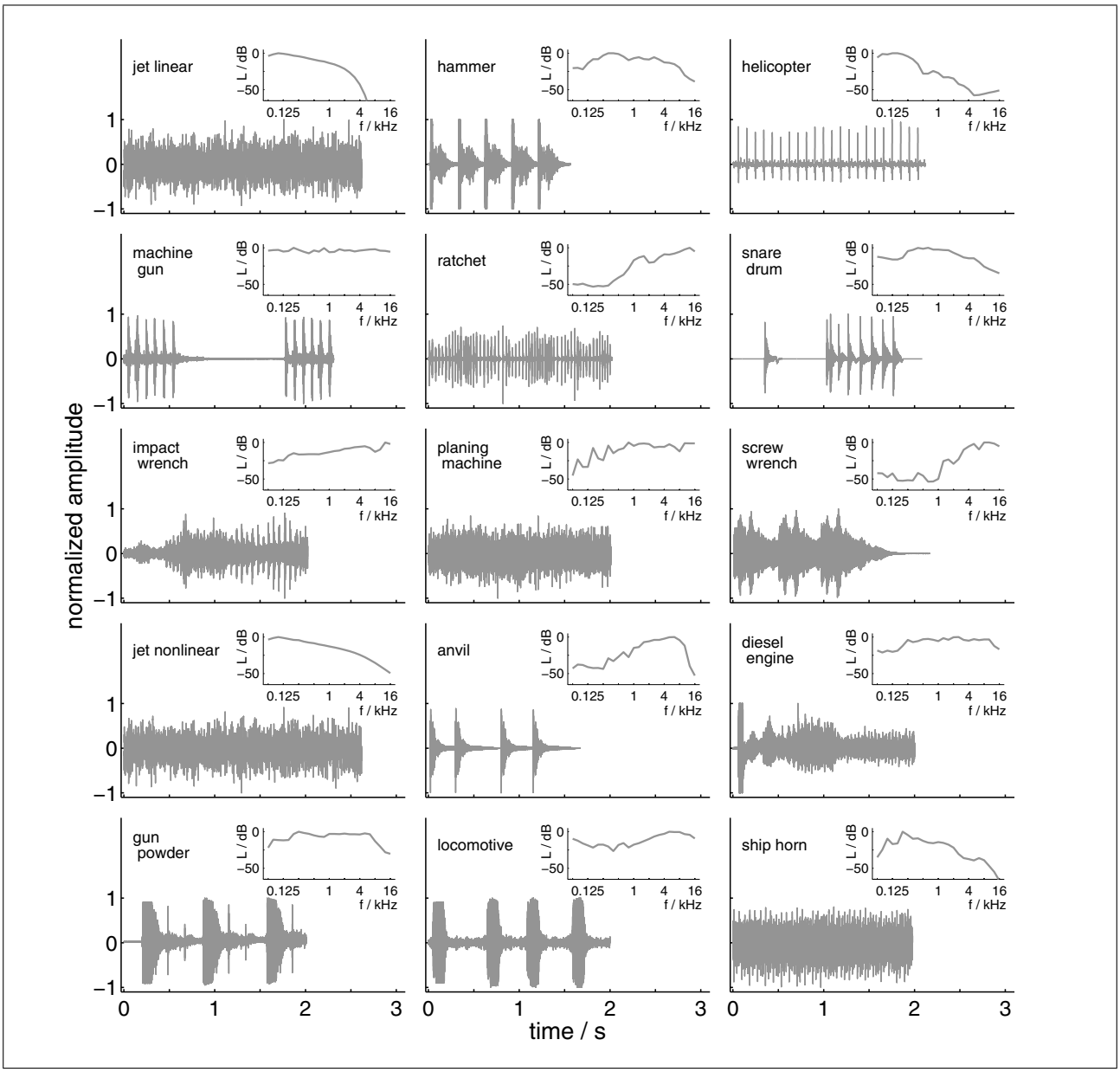

Figure 1. Waveforms of the technical sounds used in this study. The amplitudes were normalized to a maximum sample magnitude of unity. Insets represent long-term one-third-octave spectra. The signal *jet linear* (top left panel) was used as reference signal in all experiments.

loudness as well as signals for which predicted level differences differed markedly between models.

In Experiment 2, artificial stimuli were generated from the original sounds by removing original temporal envelope fluctuations, but preserving the original long-term one-third-octave spectra. This was achieved by filtering white noise using an FIR filter with 4097 coefficients to match the long-term one-third-octave spectrum to that of the original signal. The remaining spectral differences were again equalized using newly designed FIR filters until a total of ten such iterations were performed. The subset of signals was selected to include a large variety of long-term spectra.

In Experiment 3, a different set of artificial stimuli was used. These stimuli approximately preserved the temporal envelope structure of the original sounds, but had a highly similar long-term one-third-octave spectrum (namely that of the sound *jet linear*). For each test signal, this was achieved by first dividing the sound *jet linear* sample-wise by its Hilbert envelope (to create a flat temporal envelope). The resulting signal was then multiplied by the Hilbert envelope of the test signal. This signal was then subjected to the same equalization procedure of the long-term one-third-octave spectrum as used for the set of stimuli of Experiment 2 (10 iterations of FIR filter design and spectral equalization) to minimize remaining spectral differences between the signals. Essentially the same subset of signals as in Experiment 1 was selected for Experiment 3. The only exception was the signal *jet nonlinear*, which had a rather flat temporal envelope and, therefore, no large dif-

ferences in loudness compared to the reference signal were expected.

2.2. Procedure

Level differences at equal loudness were measured using an adaptive two-alternative forced-choice procedure. The levels of the test and reference signals were calculated from the root-mean-square (rms) values of the whole temporal waveforms. Subjects heard a reference and a test signal separated by 500 ms of silence and their task was to indicate which of the sounds was perceived as louder. The order of test and reference signals was randomized for each presentation. Each interval had a duration of 2.62 s corresponding to the duration of the longest stimulus of the set of stimuli used in this study. Zeros were appended to shorter test signals. The level of the reference signal was fixed at 60.8 dB SPL. This reference level corresponds to medium loudness as shown in a previous study using categorical loudness scaling [31]. The level of each equally-loud test signal was determined using an adaptive one-up one-down procedure: During an adaptive run the test-signal level was increased if the subject indicated the test signal to be softer than the reference in the previous trial; otherwise it was decreased. The initial step size was 8 dB. The step size was halved after every upper reversal until a step size of 2 dB was reached. The measurement was then continued until four further reversals were reached. The mean value of all levels which were presented after reaching the minimum step size was taken as an estimate of the level at which the test signal was perceived as equally loud to the reference signal. To avoid bias effects, each test signal was matched three times using starting levels of −10, 0, or +10 dB relative to the level corresponding to medium loudness as determined by categorical loudness scaling for the individual sounds [31]. For all three experiments, the same reference signal was used (*jet linear*). For each experiment, the different runs were divided into sessions of six runs, which were measured in an interleaved way as in, e.g., [22, 32]. Interleaved tracks were used to further reduce bias effects [33]. A test signal occurred only once in each session, and the division of test signals into sessions was randomized for each subject. Stimuli presentation was controlled using Matlab with a sampling rate of 44.1 kHz. Stimuli were D/A-converted (RME ADI-8 PRO), attenuated (Tucker Davis HB7) and presented diotically via Sennheiser HD650 headphones. All measurements were conducted in sound-attenuated booths.

2.3. Subjects

Eleven normal-hearing subjects participated in Experiment 1. The same subjects participated in Experiment 2. Twelve subjects participated in Experiment 3. Six of them had also participated in the other experiments. None of the subjects reported any hearing difficulties. All had pure tone thresholds not exceeding 15 dB hearing level (HL) at audiometric frequencies between 125 and 8000 Hz, except for one subject who had a pure tone threshold of 25 dB HL at 8 kHz in one ear. Two of the subjects were the first and second authors of this paper.

2.4. Loudness models

Eight loudness models were used to predict the experimental data. Two of the models are for stationary sounds, i.e., they account for long-term spectral effects only but cannot predict temporal effects of loudness. The other models predict loudness as a function of time and are referred to as dynamic loudness models in the following. The output of the dynamic models is loudness (in sones) as a function of time. To derive an overall loudness of a particular sound from the loudness-time-function, different criteria have been suggested (such as the maximum, mean, or a specific percentile). The models including the chosen criteria are described in more detail below.

For each model, predictions were derived using a procedure similar to that used in the experiment, i.e., the level of the test signal was varied until the predicted loudness was equal to the predicted loudness of the reference signal at the reference level. Silent intervals of 100 ms were added to the beginning and the end of the signals for the loudness calculations. The response to the silent intervals was discarded from the calculations of the mean or percentile values of the loudness-time functions. The diffuse-field equalization of the HD650 headphones used in the experiment was mimicked by pre-processing the signals using an FIR filter with 2049 coefficients. The following models were used (for details see the referenced standards and journal publications):

DIN 45631: German national standard for calculating the loudness of stationary sounds [3]. The model predicts spectral loudness summation by a bank of critical-band filters, followed by channel-wise compression and a subsequent summation across filters.

ANSI S3.4: American national standard for calculating the loudness of stationary sounds [2]. The structure of this model is similar to that of DIN 45631, but differs at various stages (e.g., filter characteristics and compression ratios) and additionally comprises a filter simulating the transmission characteristics of outer and middle ear.

DIN 45631/A1: Extension of DIN 45631 [3] to predict the loudness of time-varying signals, which applies the spectral processing of DIN 45631 to short time frames resulting in a loudness value for each frame (instantaneous loudness). This loudness-time function is smoothed using a temporal integration stage consisting of two parallel lowpass filters with different time constants [16]. The 95th percentile of the short-term loudness was used to derive overall loudness in the present study.

DLM: Dynamic loudness model proposed in [21]. The general structure is similar to DIN 45631/A1 in that a stage accounting for spectral loudness summation is applied to short time intervals and the resulting loudness-time-function is smoothed by a subsequent temporal integration stage. In the DLM, the temporal integration is

realized by a single leaky integrator (8-Hz lowpass filter). The 95th percentile of the short-term loudness was used to derive overall loudness in the present study.

eDLM: Extension of the DLM as proposed in [26]. This model introduces a bandwidth-dependent loudness increase for impulsive signals and was shown to be able to predict increased spectral loudness summation of short noise bursts [22, 23, 24, 25], which could not be predicted by independent spectral and temporal loudness integration stages [26]. This model uses the same leaky integrator as the DLM (single lowpass filter). The 95th percentile of the short-term loudness was used to derive overall loudness in the present study.

eDLMmod: Modified version of the eDLM as proposed in [14]. Instead of the single lowpass filter, two parallel lowpass filters with time constants of 45 and 185 ms are used in the temporal integration stage. The time constants of the filters were set to better predict temporal integration of short tone bursts [13, 34]. The 95th percentile of the short-term loudness was used to derive overall loudness in the present study.

stTVL: Model to predict time-varying loudness proposed in [17]. The general structure of the TVL is similar to that of the DLM, but the temporal integration stage is realized as a stage resembling automatic level control with a short attack time constant (22 ms) and a longer release time constant (50 ms). The 95th percentile of the short-term loudness was used to derive overall loudness in the present study[1].

ltTVL: In contrast to the other time-varying loudness models, the TVL comprises a second cascade of temporal integration. The short-term loudness-time-function is smoothed further by a second temporal smoothing stage with an attack time constant of 99 ms and a release time constant of 2000 ms. This long-term loudness is used to predict the loudness of longer stimuli slowly varying in loudness [17]. As proposed in [17], the mean of the long-term loudness was used in the present study to derive overall loudness.

3. Results

3.1. Experiment 1: Loudness of sounds with varying spectro-temporal characteristics

The filled circles in each panel of Figure 2 show mean level differences between the test signals and the reference signal at equal loudness for the original, unmodified sounds. Error bars indicate ±1 standard error across subjects. They ranged from 0.8 to 2.2 dB. The sounds were arranged according to their level difference at equal loudness with the lowest level difference on the far left and the highest level difference on the far right of the panel. In general, level differences were negative, i.e., a lower level was required for the test signals to be as loud as the reference signal. The minimum level difference was −13 dB (*planing machine*). A positive level difference (+7 dB) was measured only for the helicopter sound. Intraindividual variability was assessed by considering the three repetitions (i.e., the three starting levels) of each data point. While some subjects were highly consistent for all stimuli (differences in measured level differences across the three repetitions <0.5 dB), others showed larger discrepancies for some stimuli (differences up to 12 dB between the lowest and the highest measured point of subjective equality). The mean (calculated across stimuli) of the range covered by the three repetitions was 4.1 dB.

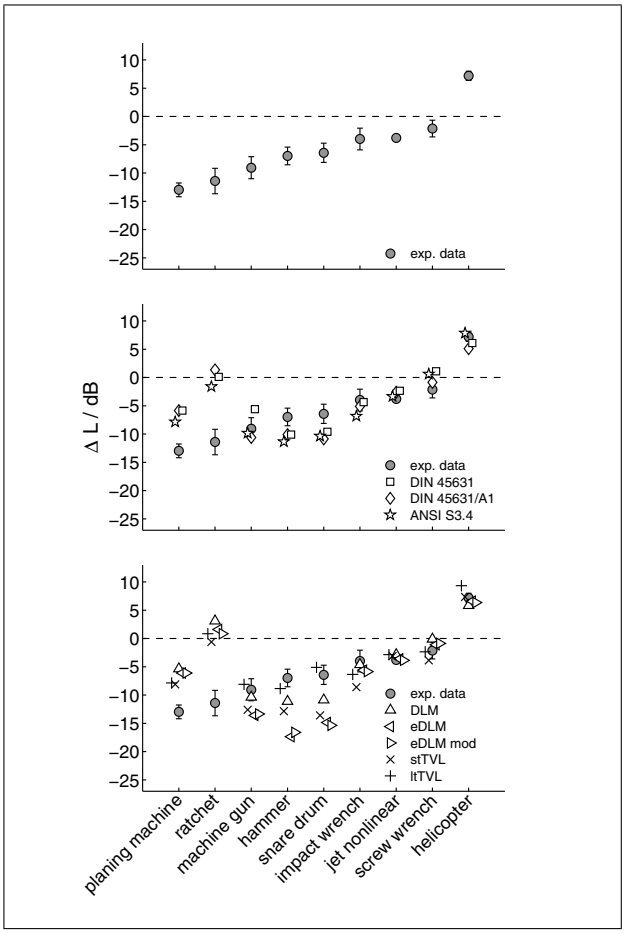

Figure 2. Mean level differences between test signal and reference signal at equal loudness for the various original technical sounds used in Experiment 1. Error bars indicate ±1 standard error of the mean. Filled circles in each panel represent experimental data. Open symbols represent data predicted by stationary (middle panel) and time-varying (middle and bottom panel) loudness models.

The middle panel shows the experimental data together with predictions of the models described in the standards for stationary sounds (open symbols). The predictions of DIN 45631 (squares) and ANSI S3.4 (diamonds) were generally similar (differences smaller than 2 dB) except for one signal (*machine gun*). For some test signals, pre-

[1] For very brief sounds, Glasberg and Moore [17] proposed to use the maximum of the short-term loudness rather than the 95th percentile. The latter was used in the present study to be comparable to the other models based on short-term loudness. A subsequent analysis revealed that using the maximum instead of the 95th percentile did not improve the prediction accuracy of the stTVL model.

Table II. Correlation analysis of measured and predicted thresholds.

		DIN 45631	ANSI S3.4	DIN 45631/A1	DLM	eDLM	eDLM mod	stTVL	ltTVL
Exp.I	R^2	0.40	0.52	0.30	0.25	0.25	0.26	0.38	0.51
	ρ_S	0.58	0.55	0.45	0.45	0.45	0.45	0.33	0.53
	ε / dB	5.0	4.4	5.3	5.9	6.8	6.6	5.4	4.6
	bias / dB	-2.1	-0.8	-1.1	-1.5	0.5	0.6	0.7	-2.0
Exp.II	R^2	0.88	0.94	0.85	0.81	0.84	0.84	0.94	0.95
	ρ_S	0.89	0.89	0.88	0.88	0.88	0.88	0.89	0.89
	ε / dB	1.9	2.2	2.0	2.3	2.2	2.2	2.4	2.4
	bias / dB	-0.4	0.8	-0.5	-0.9	-0.2	-0.2	1.1	1.1
Exp.III	R^2	0.00	0.04	0.01	0.06	0.04	0.02	0.00	0.34
	ρ_S	-0.05	0.33	0.09	0.31	0.29	0.28	0.05	0.64
	ε / dB	1.2	1.3	1.9	2.1	2.5	2.3	2.6	2.7
	bias / dB	0.1	-0.7	1.2	1.5	1.9	1.7	2.0	-2.3
Overall	R^2	0.62	0.73	0.53	0.45	0.45	0.47	0.61	0.72
	ρ_S	0.68	0.82	0.64	0.63	0.64	0.64	0.64	0.84
	ε / dB	3.2	3.0	3.5	3.8	4.4	4.2	3.7	3.4
	bias / dB	-0.8	-0.2	-0.2	-0.4	0.7	0.6	1.2	-0.9

dicted level differences were very close to the experimental data, but for two test signals substantial deviations were observed. These were the sounds *planing machine* and *ratchet*. For those sounds, the loudness of the test signal was underestimated by both models, i.e., the predicted levels at equal loudness were higher than the measured levels at equal loudness. The largest discrepancy was observed for the signal *ratchet* (about 13 dB). For the sounds *hammer* and *snare drum* the loudness was overestimated, resulting in lower predicted level differences than observed in the measured data. The predictions of DIN45631/A1 were often close to those of the corresponding stationary model (DIN45631). An exception is for the signal machine gun where the prediction of DIN45631/A1 was about 5 dB lower than that of DIN45631 and was closer to the measured data.

The bottom panel shows the predictions of the dynamic loudness models. Triangles indicate predictions of different versions of the DLM, crosses those of the two versions of the TVL. In general the differences between the predictions of the dynamic models were larger than for the stationary models. The largest differences (of up to 10 dB) were obtained for the *hammer* and the *snare drum*. For these sounds only the predictions of the ltTVL were close to the measured data. All other models overestimated the level differences by several decibels. The smallest difference between the model predictions was found for the sound *jet nonlinear*. For this sound all models predicted the measured level difference (deviation smaller than 1 dB). For some signals, level differences predicted by the dynamic models were similar to those for the stationary models. This also included the signal *ratchet*, for which predictions of dynamic models were only slightly worse than those of the stationary models. For the signals *snare drum* and *hammer*, level differences predicted by models using short time constants deviated from the data as much as or even slightly more than observed for stationary models. For these signals, ltTVL provided the best predictions of the data. In general, the model predictions employing long time constants (ltTVL) tended to be closer to the data than for the other dynamic models.

To assess the overall prediction accuracy of the models, the coefficient of determination (R^2), Spearman's rank correlation coefficient (ρ_S), the root-mean-square error (ε), and the prediction bias, i.e., the mean signed difference between measured and predicted level differences were calculated (see top rows of Table II). These measures supported the observation that predictions of stationary models and ltTVL were closest to the experimental data. The largest R^2 was observed for ANSI S3.4 (0.52) and ltTVL (0.51), while the other models did not explain more than 40% of the variance in the data. The rank correlation was similar for DIN 45631 (0.58), ANSI S3.4 (0.55), and ltTVL (0.53). Poorest prediction accuracy was observed for the different versions of the DLM.

3.2. Experiment 2: Loudness of stationary sounds with varying spectra

Figure 3 shows data (filled circles) and predictions (other symbols) for the stationary sounds with the same spectra as the technical sounds. As in Figure 2, the sounds are ordered by the measured level differences at equal loudness. Note that the sounds indicated on the abscissa are different from those used in Experiment 1. The range of measured level differences was 15 dB, i.e., slightly smaller than for the original signals. This is partly due to the fact that the original sound of Experiment 1 with the largest level difference (*planing machine*) was not part of the set of sounds used in Experiment 2. The lowest level difference of −10 dB was measured for the sound *machine gun*.

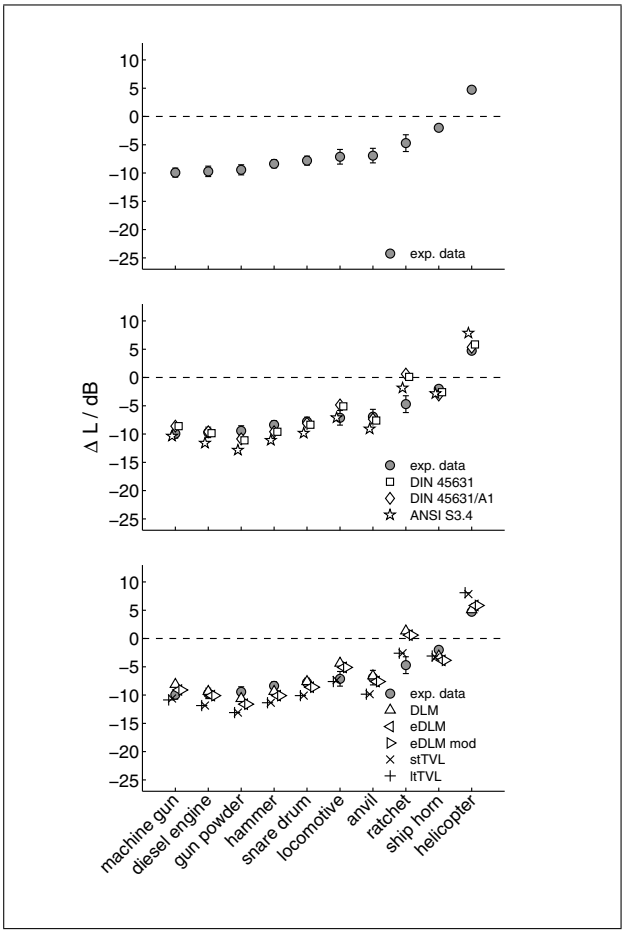

Figure 3. Same representation as Figure 2, but now showing data for Experiment 2 where artificial sounds with the same spectra as the original technical sounds were used, but with the temporal fluctuations of the original sounds largely removed.

A very similar level difference was measured for the original sound (Figure 2, third data point from the left). As in Experiment 1, the only positive level difference was found for the sound *helicopter*. The difference amounted to +5 dB, i.e., it was 2 dB smaller for the original sound. Standard errors across subjects were considerably smaller (between 0.3 and 1.5 dB) than in Experiment 1, indicating that subjects were more consistent in their judgments. This was also observed for the intraindividual variability. The mean (calculated across stimuli) covered by the three repetitions was 2.8 dB, i.e., smaller than in Experiment 1. The extremes were very similar to Experiment 1, i.e., individual subjects differed markedly across repetitions for some stimuli (up to 12 dB), while other subjects were highly consistent (differences <0.5 dB for all stimuli).

For this set of stimuli, predictions of the models described in the standards (middle panel) were generally closer to the experimental data than for the original sounds of Experiment 1. As expected for this set of signals, the predictions of DIN45631 and DIN45631/A1 were about the same (deviations less than 1 dB). In general, predictions of DIN 45631 deviated by less than 2 dB from the measured results. The only exception was for the sound ratchet, where the deviation was 4.6 dB. Deviations of level differences predicted by ANSI S3.4 were slightly larger than 2 dB for four of the test signals (*anvil*: 2.2 dB, *helicopter*: 3.4 dB, *ratchet*: 3.1 dB, and *gun powder*: 2.5 dB).

Likewise, predictions of the dynamic loudness models (bottom panel of Figure 3) were generally in good agreement with the experimental data. The models stTVL and ltTVL slightly overestimated the loudness of the sounds *diesel engine, gun powder, hammer, snare drum*, and *anvil*, and slightly underestimated the loudness of the signal *helicopter*. For these signals, predictions of DLM and its modified versions were slightly closer to experimental data than those of stTVL and ltTVL. For the signals *locomotive* and *ratchet*, predictions of the TVL model versions were closer to the measured results than the predictions of the DLM model versions. The better prediction accuracy for the stimuli of Experiment 2 was also reflected in the measures shown in Table II. All models could predict 81% (DLM) or more of the variance of the data. Highest R^2 values of 0.94 / 0.95 were observed for the two TVL versions, which had a slightly larger prediction bias (>1 dB) than the other models. All model predictions were similarly accurate in terms of rank correlations (ρ_S of 0.88 or 0.89).

3.3. Experiment 3: Loudness of time-varying sounds with matched spectrum

Mean data (filled circles) and model predictions for the artificial sounds of Experiment 3 (matched spectra, different time-varying envelopes) are shown in Figure 4. Note that the original sounds of Experiment 1 with the extreme results (*planing machine, helicopter*) were also included in this set of artificial sounds, which were derived by only preserving the temporal envelope of the original sounds (and not their spectra). However, the measured level differences only ranged from -2.5 to +1 dB, i.e., the range was much smaller than in Experiment 1. The interindividual standard errors ranged from about 0.5 dB (*impact wrench*) to about 1.5 dB (*snare drum*), i.e., they were somewhat larger than in Experiment 2 and similar to Experiment 1. The intraindividual variability was similar to that observed for Experiment 2 (and hence smaller than observed in Experiment 1). The mean (calculated across stimuli) covered by the three repetitions was 2.4 dB, respectively. As in Experiment 1, some subjects were highly consistent (differences <0.5 dB for all stimuli), while individual subjects differed markedly across repetitions for some stimuli (up to 8 dB).

Since the spectra of all sounds (including the reference) were about the same, the stationary models described in the standards ANSI S3.4 and DIN 45631 should predict level differences close to 0 dB for all sounds. In agreement with this hypothesis, the predicted level differences of the stationary loudness models (middle panel) were in a range of ±2 dB around 0 dB indicating that the matching of the long-term spectra to the reference signal had been reasonably accurate in terms of loudness equalization. The dy-

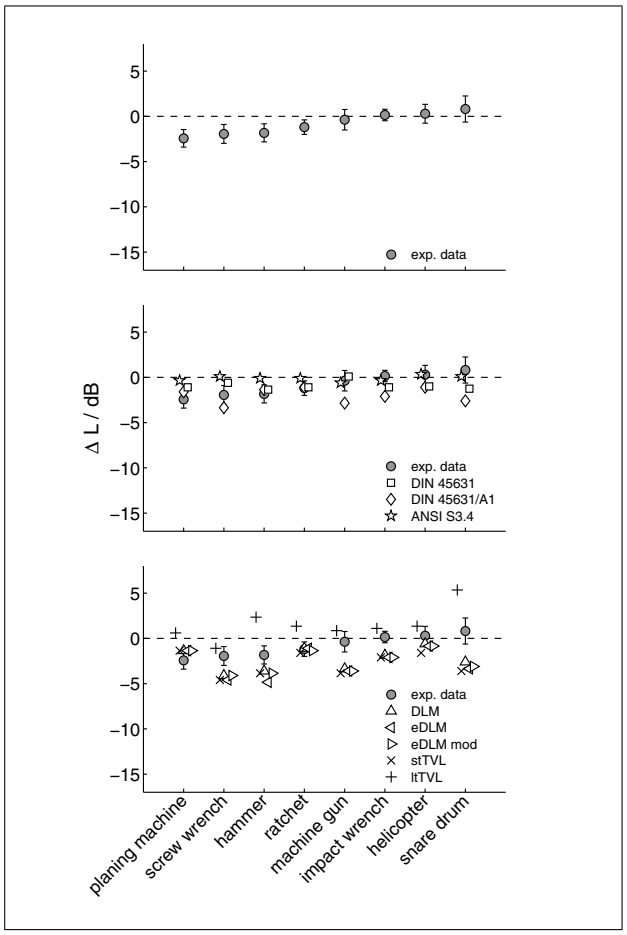

Figure 4. Same representation as Figures 2 and 3, but now showing data for experiment 3 where sounds with the same temporal envelope as the original technical sounds were used, while the spectrum for all sounds was that of the sound *jet linear*.

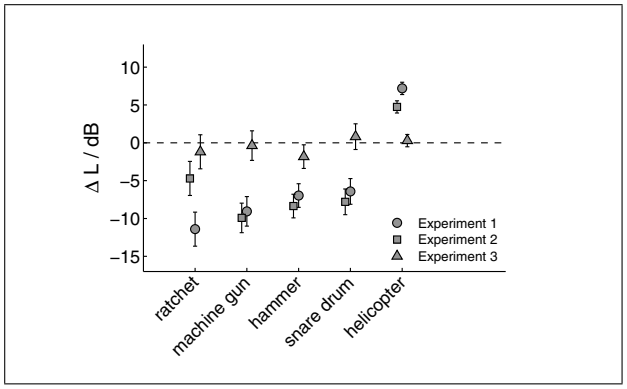

Figure 5. Level difference at equal loudness of the five sounds that were included in all three experiments. Results for the original sounds (Experiment 1) are shown by circles, those where only the spectrum was preserved (Experiment 2) by squares, and those where only the temporal envelope was preserved (Experiment 3) by triangles.

namic model described in DIN45631/A1 predicted about 3 dB lower level differences at equal loudness than the stationary model described in DIN45631 for the signals *screw wrench* and *machine gun*. For the other sounds the predictions of these two models were about the same. The level differences predicted by dynamic loudness models (bottom panel) showed larger variations than the stationary models. For most sounds (*screw wrench, hammer, machine gun, impact wrench, snare drum*), the loudness of the test signals was considerably overestimated, resulting in predicted level differences which were several decibels below those found in the experimental data. The prediction was worst for the sound snare drum, where the deviation from the measured level differences was 5 dB.

The limited prediction accuracy was also reflected in the analysis measures shown in Table II. They revealed a poor prediction accuracy of all models. R^2 values did not exceed 0.06 for stationary models as well dynamic models based on short-term loudness. The best but still poor R^2 value was observed for ltTVL (0.34).

4. Discussion

Five sounds were included in all three experiments. Their results are shown in Figure 5. Each symbol indicates one experiment (circles: Experiment 1, squares: Experiment 2, triangles: Experiment 3). For most sounds, the squares were close to the circles, i.e., results were similar for the original sound and a sound where only the spectrum was preserved. In contrast, the results of Experiment 3 were hardly influenced by the type of sound and differed considerably from those for the original technical sounds. Thus, for these sounds the spectrum seemed to dominate loudness and temporal aspects had little influence. An exception was the sound *ratchet*. Here the loudness matches for the sounds where only one of the properties was preserved (Experiments 2 and 3) differed by more than 6 dB from that for the original sound (Experiment 1). This suggests that for the sound *ratchet* it is not appropriate to assume that temporal and spectral properties are analyzed independently and that the spectrum dominates loudness.

In contrast to many previous loudness studies which used artificial sounds (e.g., tones or noise bands), the present study employed sounds from real technical devices. It is therefore possible that higher-level perceptual factors affected loudness judgements in this study, in that subjects may have tried to estimate the power or strength of the sound source rather than the power reaching the ear. For example, for a sound like ratchet, subjects may have estimated the strength with which the ratchet hits the wheel. The impact of context knowledge such as room reverberation characterizing the sound transmission path on perceptual judgements has been shown, e.g., for perceived distance [35]. Similarly, perceived annoyance was found to depend significantly on the subjects' ability to identify the sound sources [36]. Since such psychological factors are not taken into account in the tested loudness models, they might have contributed to the discrepancies between data and predictions. There are, however, two indications that such factors did not affect the subjects' judgements in the present study to a very large degree. First, the rank order of measured loudness does not correspond to the order

of source strength. For the original sounds used in Experiment 1, one can assume that subjects were able to identify the sound sources based on the acoustic stimuli at least for some very obvious signals (e.g., hammer, helicopter, machine gun, and ratchet). If their choices had been based on the source strength or source power rather than on the power reaching the ear, then one would have expected, e.g., the signal helicopter to be rated significantly louder than ratchet, or jet noise to be rated louder than snare drum. This was not observed in the data. Second, results were similar for unrecognizable sounds. When the original sounds were reduced to stationary signals with only their spectrum kept intact (Experiment 2), subjects could no longer identify the sources. This experiment showed that the signals hammer and snare drum were still judged equally compared to the reference sound.

A sound whose spectrum is likely to largely determine the loudness is the reference sound *jet linear*, which shows only subtle temporal fluctuations (Figure 1). This sound is one of the jet sounds of Gee *et al.* [29] that were included in the set of sounds used in the present study. The other sound was *jet nonlinear*, which was one of the original sounds of Experiment 1. Gee and coauthors argued that neither the traditional weighted sound pressure level nor the stationary models of the current ISO 532 could explain the clear difference in annoyance and loudness between a nonlinearly propagated jet noise and a linearly propagated jet noise, which they reported from an informal listening test. The loudness models predicted a level difference of less than 2 dB for equally loud *jet linear* and *jet nonlinear* sounds, which was, according to the authors, too small to account for the perceptual difference of the two signals. The present study measured a level difference of about 4 dB at equal loudness for these two sounds. Most models predicted level differences that were slightly smaller than the measured level difference, but the deviations were small (less than 1.5 dB). The best prediction was achieved by the modified extended dynamic loudness model (eDLM mod). This indicates that an onset accentuation as included in the eDLM models is able to account for the perceptual effect of the "staccato-like impulsive character" of these sounds, as described by Gee *et al.* [29].

Swift and Gee [37] extended the study of Gee *et al.* [29] and investigated the short-term loudness predicted by stTVL to predict the perceptual characteristics of the nonlinear jet noise. They argued that this measure cannot be used to characterize the special character of this noise. They proposed using the instantaneous loudness, which is fast enough to represent the sudden onsets in the sound *jet nonlinear*. The problem with this approach is that it is commonly assumed that this quantity is not available to conscious auditory perception (e.g., [17]). The present study indicates that it may not be necessary to use instantaneous loudness if a model working on the short-term loudness is used that includes an onset accentuation as proposed in the eDLM. While the results for the sound *jet nonlinear* might be interpreted as supporting an onset accentuation as implemented in eDLM and eDLMmod, the results for the sound *hammer* of Experiment 1 argue against this approach. Both models overestimated the loudness considerably more than the corresponding dynamic model without such an onset accentuation (DLM). For eDLM, this trend was also observed in the results for the sound *hammer* of Experiment 3, where only the temporal envelope was preserved. Interestingly, this was not true for eDLMmod, which predicted about the same level difference as the original DLM for the sound with the temporal envelope of the sound *hammer* of Experiment 3. Thus, for this sound the predictions were affected by onset accentuation and the characteristics of the temporal integration stage. However, in Experiment 3 the effects were generally small.

The sound *snare drum* of Experiment 1 is another example where the onset accentuation of eDLM and eDLMmod resulted in poorer predictions than with the corresponding model without such a stage (DLM). For this sound, hardly any differences were observed when only the temporal envelope was preserved (Experiment 3). Note that, for most sounds, the predictions of DLM, eDLM, and eDLMmod were very similar.

Figure 6 shows typical examples with similar predictions of the models stTVL, DLM, and eDLM. These models were selected since they are all based on the short-term loudness and are thus comparable with respect to their temporal processing characteristics. The upper panels show cochleograms of these sounds, which were generated by analyzing the sound with a bank of one-ERB_N wide 4th order gammatone filters [38]. The filters were spaced at one-ERB intervals. For each filter, the magnitude of the analytical output of each filter was calculated as a function of time. The magnitude-time function was then smoothed with a 1st order Butterworth filter with a cut-off frequency of 120 Hz to account for the limited ability of the auditory system to follow amplitude modulations. The choice of the cut-off frequency was motivated by previous studies on modulation-detection thresholds. They indicated low-pass filtering with a cut-off frequency in the range from 50 to 150 Hz [39, 40]. The magnitude is indicated by gray values with black representing very large magnitudes. The lower panels show the differences between predicted and measured levels at equal loudness. For the sound *ratchet* (left panel), all models predicted level differences that were larger than measured, i.e., the models underestimated the loudness of the sound relative to the reference sound. For the sound *snare drum* (middle panel), the models overestimated the loudness. For the sound *helicopter* (right panel), all models predicted the measured level difference. The cochleograms show that the spectral decomposition at the level of the auditory periphery preserves the difference in spectral center of gravity. However, these are unlikely to account for the difference in accuracy of the predictions, since the model predictions were close to the data of Experiment 1 for the signal *screw wrench*, which had a similar spectrum to that of the *ratchet*. Thus, the interaction of temporal structure and spectrum is likely to be the reason for the difference in accuracy of the model predictions.

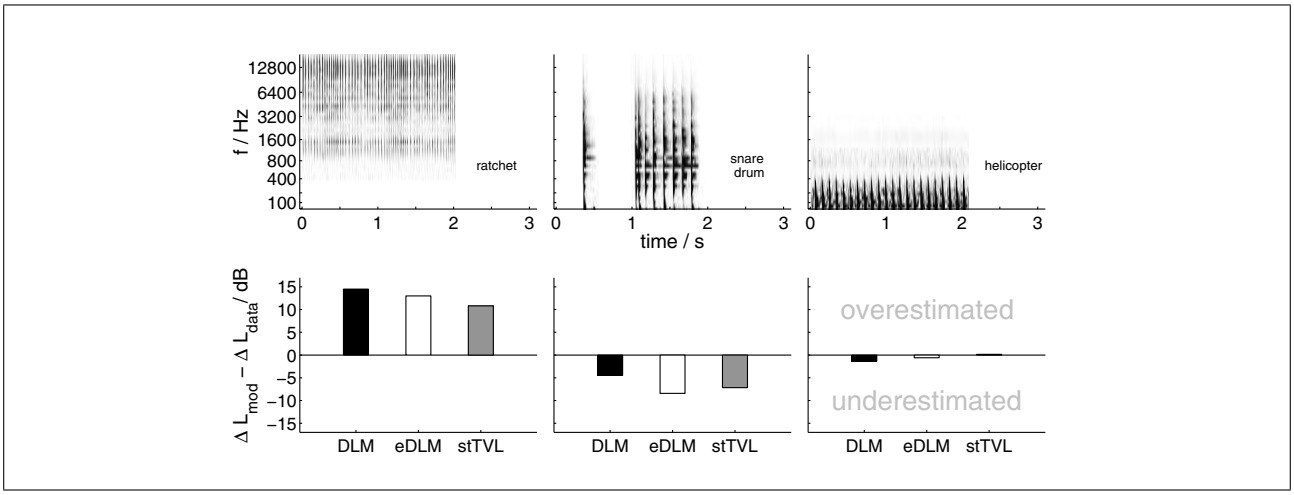

Figure 6. Top row: Cochleograms of the sounds *ratchet* (left), *snare drum* (middle), and *helicopter* (right). Bottom row: Differences between predicted and measured level differences at equal loudness for the models DLM, eDLM, and stTVL. Positive values indicate that the level difference between equally loud test and reference signal were overestimated by the models.

This becomes obvious for the signal *snare drum* for which all models shown in Figure 6 underestimated the loudness. In contrast, ltTVL predicted the data (see Figure 2). Interestingly, the predictions of the stationary models are not as good as those of ltTVL. This indicates that the long decay time constant may be crucial for this particular signal since it fills in the gap between the first beat and the following beats. For the helicopter, where such a long temporal gap does not occur, the predictions of the three models shown in Figure 6 were much closer to the data. For the *ratchet*, the temporal fluctuations were considerably faster than those for the other two signals. This temporal feature in combination with a high spectral center of gravity may have led to the overestimation of the loudness for this sound.

The smallest deviations between predictions of stTVL and the versions of the DLM and DIN 45631/A1 were found for the temporally fluctuating sounds of Experiment 3. This is consistent with Rennies and colleagues, who showed that stTVL and DLM can predict loudness for a number of artificial time-varying stimuli with comparable accuracy [41]. Differences between the predictions of stTVL and DLM occurred for the original technical sounds of Experiment 1 and the stationary sounds with original spectra of Experiment 2. There was no clear trend with respect to the accuracy of the predictions of the data of Experiment 1. For some sounds stTVL was closer to the measured data (e.g., *planing machine, ratchet*) while, for other sounds, DLM was more accurate (e.g., *machine gun, impact wrench*). A preference for one of the dynamic models DLM, eDLM, eDLMmod and stTVL could not be derived from Experiment 2 either. Since temporal aspects should not influence the results of Experiment 2, the difference between stTVL and the models based on the DLM can be attributed to differences in model parameters such as filter width and frequency weighting. DIN45631, DIN45631/A1 and the various versions of the DLM all predicted about the same level differences, since they all used the same frequency weighting and definition of filter width (in Bark). The stTVL model predicted the same level differences as ltTVL and ANSI S3.4, where the same frequency weighting and filter widths (in ERB_N) were used, which differed from those of the other models. In contrast to [11] the present study does not favor DIN45631 over ANSI S3.4.

As expected, the predictions of stTVL and ltTVL were different when temporal aspects of the sounds played a role, as in Experiments 1 and 3. For most sounds of Experiment 1, the predictions of ltTVL were closer to the data than those of stTVL. Only for the sound *helicopter* and *ratchet* were the predictions of stTVL a few decibels closer to the measured data than the level differences predicted by ltTVL. For Experiment 3, the results were less clear. For the sounds *screw wrench, machine gun, impact wrench*, and *helicopter*, the predictions of ltTVL were closer to the data than those of stTVL. The opposite was true for the sounds *planing machine, hammer*, and *ratchet*. For *snare drum*, the loudness was underestimated by several decibels by ltTVL, whereas it was overestimated by a similar amount by stTVL.

Overall, the dynamic loudness model with the longest time constants (i.e., ltTVL) was better than models with shorter time constants for most of the sounds. The latter models tended to overestimate the loudness of the temporally modulated sounds. These findings are consistent with those of a recent study of Rennies and colleagues [42]. They found that for speech and speech-like signals the long-term loudness of TVL was a better predictor of loudness than the short-term loudness of TVL and DLM. The results of the present study could therefore indicate that, for most sounds, the long-term spectrum dominated the loudness. The sounds *ratchet* and *planing machine* of Experiment 1 were poorly predicted by all models used in the present study. The experiments with signals where only the temporal or the spectral characteristics were preserved seemed to point towards a possible explanation. Figure 5

shows that, among the five sounds that were used in all three experiments, the *ratchet* was the only sound where the spectrum did not dominate overall loudness. Thus, one might argue that all signals with this characteristic were not predicted by current loudness models. Further experiments are required to test this hypothesis.

5. Conclusions

- The loudness of technical sounds is often dominated by their spectral characteristics rather than by their temporal properties. The loudness of stationary sounds can be accurately predicted by current standards and loudness models.
- In general, the longer time constants of the ltTVL model yield better loudness predictions for technical sounds than shorter time constants which are commonly used for the derivation of short-term loudness.
- For some technical sounds, the loudness could not be predicted by any of the tested models.
- A particularly large discrepancy between measured and predicted loudness was observed for a signal for which the loudness was neither dominated by the long-term spectrum nor by the temporal properties. This indicates that complex spectro- temporal effects in loudness are not fully accounted for by current loudness models.

Acknowledgement

This study was partly supported by the Deutsche Forschungsgemeinschaft (SFB trr31) and the Ministry for Culture and Science of Lower Saxony, Germany.

References

[1] B. Berglund, U. Berglund, T. Lindvall: Scaling loudness, noisiness and annoyance of community noises. Journal of the Acoustical Society of America **60** (1976) 1119–1125.

[2] ANSI S3.4-2007: Procedure for the computation of loudness of steady sounds. American National Standards Institute, Washington, DC, USA, 2007.

[3] DIN 45631: Procedure for calculating loudness level and loudness. Deutsches Institut für Normung, Berlin, Germany, 1991.

[4] ISO 532:1975: Acoustics - Method for calculating loudness level. International Standardization Organization, Geneva, Switzerland, 2012.

[5] ISO 226: Acoustics - Normal equal loudness-level contours. International Organization for Standardization, Geneva, Switzerland, 2003.

[6] W. Heeren, V. Hohmann, J. E. Appell, J. L. Verhey: Relation between loudness in categorical units and loudness in phons and sones. Journal of the Acoustical Society of America **133** (2013) EL314–EL319.

[7] E. Zwicker: Subdivision of the audible frequency range into critical bands (Frequenzgruppen). Journal of the Acoustical Society of America **33** (1961) 248.

[8] B. R. Glasberg, B. C. J. Moore: Derivation of auditory filter shapes from notched-noise data. Hearing Research **47** (1990) 103–138.

[9] C. Jurado, B. C. J. Moore: Frequency selectivity for frequencies below 100 Hz: comparisons with mid-frequencies. Journal of the Acoustical Society of America **128** (2010) 3585–3596.

[10] C. Jurado, C. S. Pedersen, B. C. J. Moore: Psychophysical tuning curves for frequencies below 100 Hz. Journal of the Acoustical Society of America **129** (2011) 3166–3180.

[11] J. Schlittenlacher, W. Ellermeier: Spectral loudness summation: Shortcomings of current standards. Journal of the Acoustical Society of America **137** (2015) EL26–EL31.

[12] J. L. Verhey: Temporal resolution and temporal integration. – In: The Oxford Handbook of Auditory Science – Hearing. Volume 3. C. J. Plack (ed.). Oxford University Press, Oxford, 2010, 105–121.

[13] T. Poulsen: Loudness of tone pulses in a free field. Journal of the Acoustical Society of America **69** (1981) 1786–1790.

[14] J. Hots, J. Rennies, J. L. Verhey: Modeling temporal integration of loudness. Acta Acustica united with Acustica **100** (2014) 184–187.

[15] E. Port: Zur Lautstärkeempfindung und Lautstärkemessung von pulsierenden Geräuschen (On the perception and measurement of loudness of pulsating sounds). Acustica **13** (1963) 224–233.

[16] R. Sottek: Progresses in standardizing loudness of timevariant sounds. AIA-DAGA 2013, Merano, Italy, 2013, 1125–1128.

[17] B. R. Glasberg, B. C. J. Moore: A model of loudness applicable to time-varying sounds. Journal of the Audio Engineering Society **50** (2002) 331–341.

[18] C. Zhang, F.-G. Zeng: Loudness of dynamic stimuli in acoustic and electric hearing. Journal of the Acoustical Society of America **102** (1997) 2925–2934.

[19] G. Grimm, V. Hohmann, J. L. Verhey: Loudness of fluctuating sounds. Acta Acustica united with Acustica **88** (2002) 359–368.

[20] B. C. J. Moore, D. A. Vickers, T. Baer, S. Launer: Factors affecting the loudness of modulated sounds. Journal of the Acoustical Society of America **105** (1999) 2757–2772.

[21] J. Chalupper, H. Fastl: Dynamic loudness model (DLM) for normal and hearing-impaired listeners. Acta Acustica united with Acustica **88** (2002) 378–386.

[22] J. L. Verhey, B. Kollmeier: Spectral loudness summation as a function of duration. Journal of the Acoustical Society of America **111** (2002) 1349–1358.

[23] A.-K. Anweiler, J. L. Verhey: Spectral loudness summation for short and long signals as a function of level. Journal of the Acoustical Society of America **119** (2006) 2919–2928.

[24] J. L. Verhey, M. Uhlemann: Spectral loudness summation for sequences of short noise bursts. Journal of the Acoustical Society of America **123** (2008) 925–934.

[25] W. Heeren, J. Rennies, J. L. Verhey: Spectral loudness summation of non-simultaneous tone pulses. Journal of the Acoustical Society of America **130** (2011) 3905–3915.

[26] J. Rennies, J. L. Verhey, J. Chalupper, H. Fastl: Modeling temporal effects of spectral loudness summation. Acta Acustica united with Acustica **95** (2009) 1112–1122.

[27] W. Brennecke, H. Remmers: Physikalische Parameter bei der Bewertung der Lästigkeit von Industriegeräuschen (Physical parameters for assessing annoyance of industrial sounds). Acustica **52** (1983) 279–289.

[28] K. Johannsen, H. Prante: Environmental sounds for psychoacoustic testing. Acta Acustica united with Acustica **87** (2001) 290–293.

[29] K. Gee, S. Swift, V. Sparrow, K. Plotkin, J. Downing: On the potential limitations of conventional sound metrics

in quantifying perception of nonlinearly propagated noise. Journal of the Acoustical Society of America **128** (2007) EL1–EL7.

[30] https://www.freesound.org.

[31] M. Wächtler, I. Holube, J. L. Verhey, J. Rennies: Wahrnehmung und Modellierung von Lautheit instationärer, technischer Signale (Perception and modeling of loudness of non-stationary technical signals). Fortschritte der Akustik - DAGA, 2012, 593–594.

[32] M. Florentine, S. Buus, T. Poulsen: Temporal integration of loudness as a function of level. Journal of the Acoustical Society of America **99** (1996) 1633–1644.

[33] J. L. Verhey: Psychoacoustics of spectro-temporal effects in masking and loudness perception. BIS-Verlag, Oldenburg, ISBN 3-8142-0662-2, 1999.

[34] O. Pedersen, P. Lyregaard, T. Poulsen: The round robin test on impulsive noise. Report No 22, The Acoustics Laboratory, Technical University of Denmark, 1977.

[35] D. H. Mershon, J. N. Bowers: Absolute and relative cues for the auditory perception of egocentric distance. Perception **8** (1979) 311–322.

[36] W. Ellermeier, A. Zeitel, H. Fastl: Predicting annoyance judgments from psychoacoustic metrics: Identifiable vs. neutralized sounds. International Congress on Noise Control Engineering "Inter-Noise 2004", Prague, Czech Republic, CD–ROM (6 pages).

[37] S. Swift, K. Gee: Examining the use of a time-varying loudness algorithm for quantifying characteristics of nonlinearly propagated noise (L). Journal of the Acoustical Society of America **129** (2011) 2753–2756.

[38] V. Hohmann: Frequency analysis and synthesis using a gammatone filterbank. Acta Acustica united with Acustica **88** (2002) 433–442.

[39] N. F. Viemeister: Temporal modulation transfer functions based upon modulation thresholds. Journal of the Acoustical Society of America **66** (1979) 1364–1380.

[40] A. Kohlrausch, R. Fassel, T. Dau: The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers. Journal of the Acoustical Society of America **108** (2000) 723–734.

[41] J. Rennies, J. L. Verhey, H. Fastl: Comparison of loudness models for time-varying sounds. Acta Acustica united with Acustica **96** (2010) 383–396.

[42] J. Rennies, I. Holube, J. L. Verhey: Loudness of speech and speech-like signals. Acta Acustica united with Acustica **99** (2013) 268–282.