Loudness perception and modeling of impulsive sounds

Roland Sottek, Tobias Moll
HEAD acoustics GmbH, Herzogenrath, Germany

Summary
There are different loudness calculation procedures, such as the German standard DIN 45631/A1 and the proposed international standard ISO 532-1 as well as the Dynamic Loudness Model (DLM) (by Chalupper and Fastl), the Time Varying Loudness (TVL) model (by Glasberg and Moore), and the loudness calculation algorithm based on a hearing model of Sottek, allowing for the prediction of the perceived loudness of time-varying sounds in many cases. However, recent studies show that the predictions for some impulsive sounds do not match the ratings of normal-hearing subjects. Therefore, the influence of specific signal properties of the sounds on the assessment of loudness was examined focusing on the impulsiveness of the sounds. On the basis of these experiments, it was studied to what extent the loudness model based on the hearing model of Sottek must be adjusted to take into account the specific signal properties of impulsive sounds.

PACS no. 43.66.Ba, 43.66.Cb

1. Introduction

Loudness is one of the most important topics in psychoacoustics. Many other psychoacoustic parameters such as sharpness, roughness, tonality, and impulsiveness are related to loudness. Research on loudness has been performed for a long time with respect to experimental studies and modeling. In general, models of stationary loudness perform well for stationary signals, whereas models of time-varying loudness are suitable only to a limited extent. Recent studies of Wächtler [1] and Rennies et al. [2] showed that the predictions of loudness models for tonal and some impulsive sounds do not match the ratings of normal-hearing subjects. The challenge of loudness prediction of tonal sounds has already been discussed by Sottek [3], Hots et al. [4], [5]. This paper focuses on loudness of impulsive sounds. Figure 1 shows some loudness matching results from Rennies et al. [2] compared to predictions of various time-varying loudness models: the Dynamic Loudness Model (DLM) by Chalupper and Fastl [6], its extension by Rennies et al. (extDLM) [7], and the Time Varying Loudness (TVL) model by Glasberg and Moore using different time constants (short by default and long) [8].

There are also standards for time-varying loudness available: the German standard DIN 45631/A1 [9], which is the basis for the new ISO standard for loudness of arbitrary sounds ISO 532-1 [10], replacing ISO 532:1975 section 2 (method B) [11]. The results of ISO 532-1 and DIN 45631/A1 should be very similar, however, ISO 532-1 defines the loudness calculation procedure precisely, providing a reference implementation, whereas DIN 45631/A1 is partly imprecise and provides only few loudness results for synthetic tone pulses for the verification of an implementation. Thus, the reference implementation of ISO 532-1 is preferred over DIN 45631/A1. Results of the new ISO 532-1 standard are discussed in section 2.

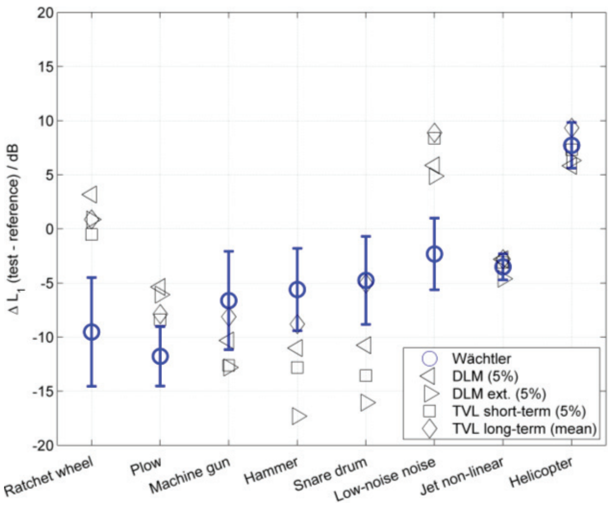

Figure 1: Results of loudness matching experiments using the indicated time-varying signals and a stationary noise as reference (blue symbols: mean and confidence interval (95 %)): the L_1[1] level differences at the point of subjective equality are given. Additionally, the predictions of various loudness models (based on N_5-values[2]) are depicted. Data is taken from Wächtler [1] (see also Rennies et al. [2]).

[1] L_1: level exceeded 1 % of the time, calculated based on the magnitude of the time signal (without any smoothing).

[2] N_5: loudness exceeded 5 % of the time; other percentile values led to similar results [1].

Copyright© (2015) by EAA-NAG-ABAV, ISSN 2226-5147
All rights reserved

None of the employed models could predict all level differences at equal loudness; especially for the ratchet wheel sound there was a discrepancy of about 12 dB. Therefore, a new study was motivated to assess the influence of specific signal properties of impulsive sounds on loudness, starting with the ratchet wheel and other sounds provided by courtesy of Rennies.

In a first step, some of the loudness matching experiments from Wächtler and Rennies, respectively, have been repeated by Parvizian [12] to verify previous experimental results.

Then in a second step, signal properties of the specific ratchet wheel sound were modified systematically (time structure, spectral content or both) in order to find a physical cause why this sound was underestimated in loudness contrary to other impulsive sounds.

After that, the hearing model of Sottek was applied to the sounds in order to find features which can be used for an improved loudness prediction of impulsive sounds.

At the end, the proposed model optimization was verified with new time data.

2. Verification of previous loudness matching results

Loudness matching was performed using an adaptive two-alternative forced-choice procedure (AFC-toolbox provided by Ewert [13]). 15 subjects heard a reference signal and a test signal separated by 500 ms of silence, presented via digital equalizer and Sennheiser HD 650 headphones. The subjects were asked to indicate which sound was perceived as louder. The sound pressure level of the test signal was adjusted depending on the subjects' response. The reference signal was a stationary noise signal 'Jet linear' with a fixed level of 61 dB and a corresponding L_1-value of 69 dB. The 8 test signals were: 'Ratchet wheel', 'Plow', 'Machine gun', 'Hammer', 'Snare drum' 'Low-noise noise', 'Jet non-linear' and 'Helicopter'. Signal durations were between 1.6 s and 2.6 s.

The matching was stopped when the results had converged to almost the same perceived loudness of the test and the reference signal.

To avoid bias effects, the test was performed three times for each test signal using different starting values: -10, 0, +10 dB relative to the starting value used by Wächtler (determined by a categorical loudness test) [1] (Table 2, first column).

The 24 runs were divided into four sessions of 6 runs each, which were measured in an interleaved way to further reduce bias effects [14].

The different starting points led to almost the same results, i.e., when a signal was tested with three different starting values, each subject came to almost the same sound pressure level and same perceived loudness, respectively. Table 1 shows an example for one subject.

Table 1: Different starting levels led to almost the same results (example for one subject).

Name	First test L_1 / dB	Second Test L_1 / dB	Third test L_1 / dB	Average L_1 / dB
Ratchet wheel	57.19	58.69	59.69	58.52
Plow	61.71	63.21	62.21	62.38
Machine gun	66.59	68.09	67.09	67.26
Hammer	74.28	72.28	74.28	73.61
Snare drum	67.74	68.24	69.74	68.57
Low-noise noise	64.85	67.35	66.35	66.18
Jet non-linear	66.62	65.12	67.12	66.29
Helicopter	81.76	80.26	81.76	81.26

For each subject and each test signal, the three results using different starting values were averaged. Table 2 shows the mean values at equal loudness for the 8 test signals based on the 11 subjects participating in the study of Wächtler and based on the 15 subjects participating in the study of Parvizian.

Table 2: Starting levels L_1 for the test signals determined by a categorical test [1] and mean values resulting from the AFC-procedures performed by Wächtler (results 1) [1] and performed by Parvizian (results 2) [12].

Name	Starting levels L_1 / dB	Results 1 L_1 / dB	Results 2 L_1 / dB
Ratchet wheel	73.69	59.47	60.71
Plow	65.21	57.22	56.21
Machine gun	71.09	62.38	65.22
Hammer	71.28	**63.39**	**68.61**
Snare drum	71.74	**64.24**	**69.00**
Low-noise noise	65.85	66.67	65.85
Jet non-linear	70.12	65.49	64.60
Helicopter	82.76	**76.72**	**83.08**

In general, the results of the two studies are fairly similar with respect to mean values and confidence intervals (95%). Larger discrepancies in the order of 5 dB can be detected for the signals 'Hammer', 'Snare drum' and 'Helicopter' for which the results of Parvizian were closer to the results of the categorical tests performed by

Wächtler. Figure 2 compares both results more in detail. Additionally, the prediction results of the new ISO 532-1 standard are shown.

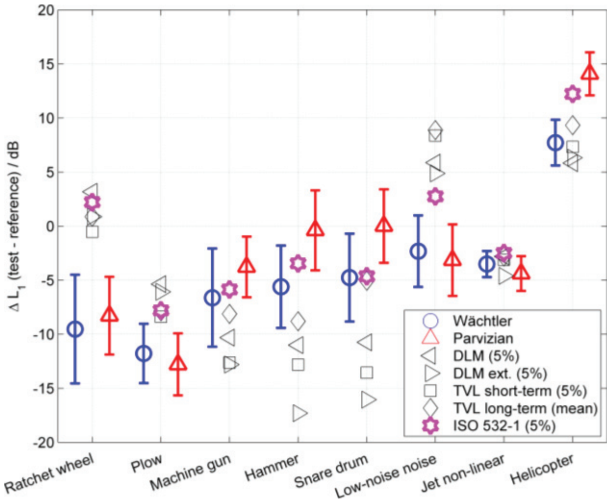

Figure 2: Results of loudness matching experiments using the indicated time-varying signals and a stationary noise as reference (blue and red symbols, mean and confidence intervals (95 %)): the L_1 level differences at the point of subjective equality are given. In addition to Figure 1, the predictions of ISO 532-1 (based on N_5-values; other percentile values lead to similar results) are given.

The prediction results of ISO 532-1 are in general closer to the results of the experiments than the other models. Nevertheless, for the ratchet wheel sound the larger discrepancy with respect to the predicted level at equal loudness, on the order of 12 dB, applies for all considered models including the ISO 532-1 standard. Moreover, all models underestimate the loudness of the plow signal and the low-noise noise. The error of the level prediction at equal loudness amounts to at least 4-5 dB (in the case of ISO 532-1).

3. Modification of signal properties

Because the loudness predictions for some impulsive sounds do not match the ratings of normal-hearing subjects, the influence of specific signal properties of these sounds on the assessment of loudness was examined. First, the time structure of the ratchet wheel sound was modified in different ways such as time reversal of single impulses or randomized order of all single impulses – without significant effects on perceived loudness. Then, new signals were generated by segmenting the signal into smaller blocks and combining them in a randomized order. The size of the blocks was varied: 500, 250, 100, 50, 25 samples and 1 sample (sampling rate: 44.1 kHz). The sounds were used together with the original ratchet wheel sound in a full paired comparison experiment performed by 17 subjects with self-reported normal hearing. All sounds were presented in a randomized order separated by 500 ms of silence. The subjects were asked to indicate which sound was perceived as louder. Only minor effects of block size on loudness could be observed. But the completely randomized signal (block size: 1 sample) was always considered as softer and the original signal as louder. In the direct comparison of both signals, 73.5 % indicated the original sound as louder in contrast to the results of the model prediction using ISO 532-1. Figure 3 shows the loudness vs. time functions of both signals whereby the curve of the randomized version is above the curve of the original sound at any given time.

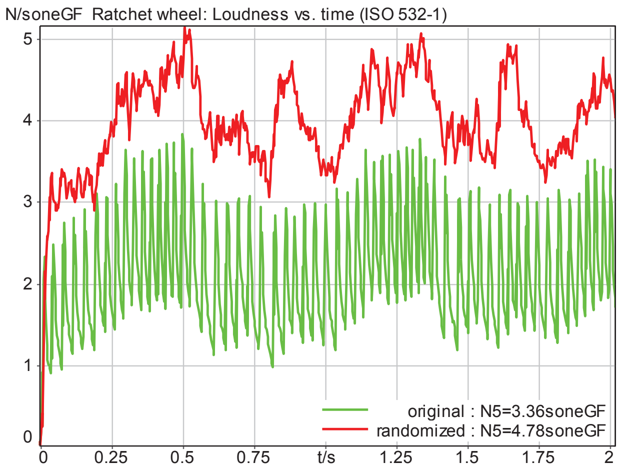

Figure 3: Results of loudness predictions using ISO 532-1 for the ratchet wheel sound and the completely randomized sound. By definition, the levels of both signals are the same: L=47.69 dB, but also their A-weighted levels differ only by 0.14 dB.

In another experiment, the third octave spectra of three sounds, a ratchet wheel sound, a hammer sound, and white noise (WN), all adjusted to the level of 47.7 dB, were adapted to the third octave spectra of the two other sounds while keeping the time structure (by means of filtering) leading to three groups of three signals, e.g., ratchet wheel sound (original), hammer sound (filtered, third octave spectrum of the ratchet wheel sound), white noise (filtered, third octave spectrum of the ratchet wheel sound) etc. These sounds were used for a loudness matching experiment using an adaptive two-alternative forced-choice procedure as described above. The original sound was always used as reference, thus leading to 6 runs which were measured in an interleaved way to further reduce bias effects. Only the original starting level was used. 15 subjects participated in this experiment.

The results of the experiment (Figure 4) show that ISO 532-1 underestimates the loudness of the ratchet wheel sounds with modified third octave spectra (L_T Hammer, L_T WN): error of the level prediction at equal loudness amounts to 3-4 dB (4-8 dB for the original ratchet wheel sound as reference stimulus, hammer and WN with third octave spectra of the ratchet wheel sound as test signals). Note, the different signs in Figure 4 for the deviation between test and prediction depend on the reference signal. Noteworthy, the N_5-values based on ISO 532-1 are almost the same for all signals with the same third octave spectra L_T, despite their different time structures (Table 3): according to ISO 532-1 the different time structures do not have a strong effect on loudness perception, which is contrary to the experimental data.

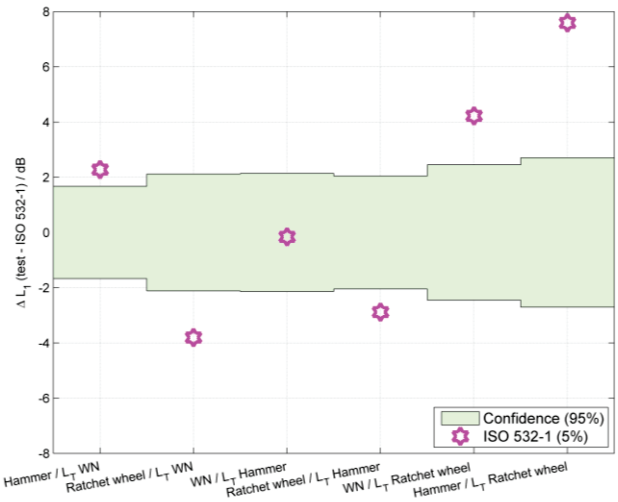

Figure 4: Differences of the results of a loudness matching experiment and the results of ISO 532-1 using filtered signals (adapted third octave spectra). The first name indicates the original signal and the second name the adapted third octave spectrum. Reference is always the original signal corresponding to the third octave level (for example WN for the first two results). Addionally, the confidence intervals (95 %) of the experiments are depicted.

Table 3: N_5-values calculated using ISO 532-1 of the nine test signals. Results for the reference stimuli are marked in bold.

Name	L_T Ratchet wheel N_5 / soneGF	L_T Hammer N_5 / soneGF	L_T WN N_5 / soneGF
Ratchet wheel	**3.44**	5.00	4,36
Hammer	3.56	**4.96**	4.53
WN	3.53	5.02	**4.42**

4. Hearing model of Sottek

Inspired by the work of Licklider [15] concerning human pitch perception, the sliding autocorrelation function has been used as a processing block in the hearing model of Sottek for the calculation of roughness and fluctuation strength [16] and later also for other psychoacoustic quantities like tonality [17]. Figure 5 displays the basic model structure.

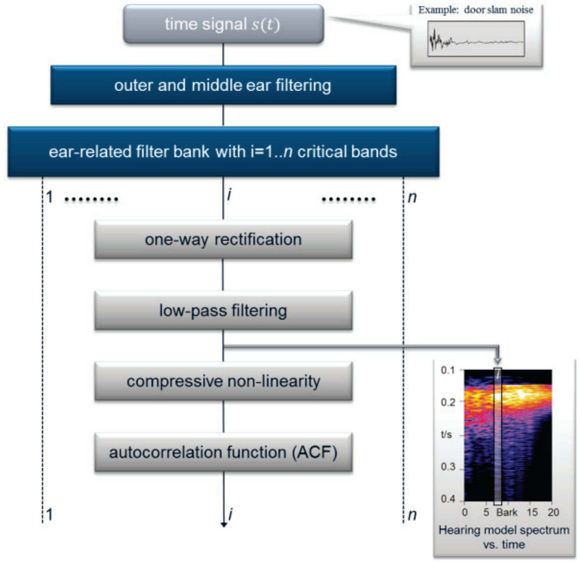

Figure 5: Basic model structure for the determination of the hearing model spectrum vs. time and the autocorrelation function as function of time, lag, and frequency band (see also Figure 6).

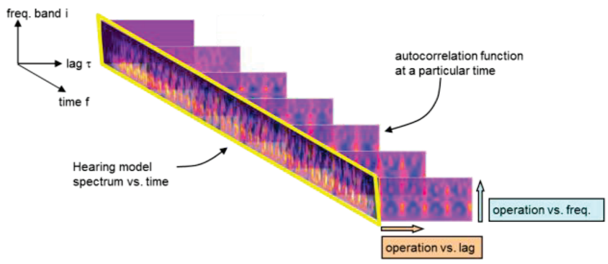

Figure 6: Sliding autocorrelation function based on the hearing model spectrum vs. time after applying a compressive nonlinearity. Different operations can be applied vs. lag and vs. frequency bands for further processing.

The many existing hearing models differ mainly in three points:

1. the frequency weighting, which is the main cause for differences in modeling equal loudness contours (especially at low frequencies: modeling the outer and middle ear transfer function, the input signal $s(t)$ is filtered by a filter $h_{am}(t)$ corresponding to the equal loudness contour at 100 phon),
2. the frequency scale (Bark or ERB) meaning the frequency-dependent bandwidth of the implemented n-channel filter bank (in this model to decompose the input signal into n critical bands, the envelope of each sub-band signal is calculated by one-way rectification),
3. the nonlinear relation between sound pressure and specific loudness (a strongly compressive

function in combination with the calculation of the autocorrelation function in each sub-band).

The nonlinearity of this hearing model uses power functions with different exponents for different level ranges [16], [18]. Such a nonlinearity function has proven applicable to predict many phenomena like ratio loudness, just-noticeable amplitude differences, and modulation thresholds as well as the level dependence of roughness.

In the first hearing model loudness approach, the calculation was based only on a summation of compressed bandpass signals [16]. Recently, it was proposed to use the autocorrelation function (ACF) of the bandpass signals to separate tonal content from noise in order to weight the loudness of tonal components differently for an improved loudness prediction of tonal sounds [19]. In the following, the ACF is used to detect special signal properties that could be used for an improved loudness prediction of impulsive sounds.

The three-dimensional ACF analysis contains a lot of information and there are many ways to analyze the data. In a first attempt, the data were condensed by summing the autocorrelation functions vs. frequency bands: over a certain number of critical bands or over all critical bands.

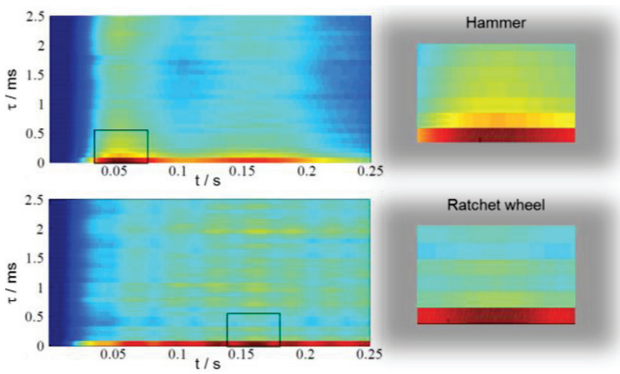

Figure 7: Sum of the autocorrelation functions of all criticial bands as a function of time t and lag τ for the hammer signal (upper) and the ratchet wheel signal (lower). The zoomed areas at the right side better show the different decays of the ACF vs. τ for the two signals.

Figure 7 shows that the decay of the ACF is much steeper for the ratchet wheel signal than for the hammer signal. Additonally, the maximum of the normalized ACF (normalization factor: ACF(τ=0)) in a certain lag range is lower for the ratchet wheel sound. Parameters related to specific frequency ranges only (low, middle, and high frequencies) could be used as further features for a recognition and an improved loudness evaluation of such ‘critical’ impulsive signals.

5. Experiments with new test signals

After observing the fact that the summed ACF of the ratchet wheel sound has different features than the hammer signal and all the other sounds used in the described experiments new test signals were identified with similar properties. As examples, another ratchet wheel sound ‘Ratchet wheel 2’, a ‘Spray can’ rattling sound, and a ‘Door creak’ sound were analyzed.

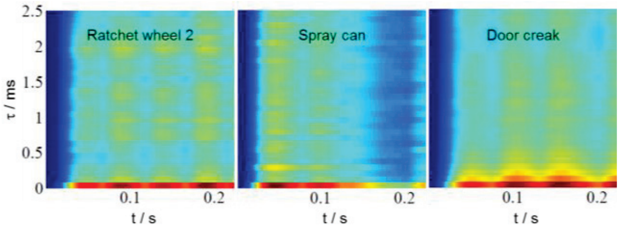

Figure 8: Sum of the autocorrelation functions of all criticial bands as a function of time t and lag τ for the ‘Ratchet wheel 2’ signal, the ‘Spray can’ rattling sound and the ‘Door creak’ sound.

On the one hand, Figure 8 shows that the results for the ‘Ratchet wheel 2’ and the ‘Spray can’ rattling sounds have one key aspect in common with the result of the first ratchet wheel sound: an abrupt decay of the normalized summed ACF. These sounds are very impulsive and have strong high frequency contents. On the other hand, the decay of the normalized summed ACF of the door creak sound is very similar to that of the hammer sound. These sounds are less impulsive and have stronger low frequency contents.

All 5 sounds were used for another loudness matching experiment using an adaptive two-alternative forced-choice procedure. The stationary sound ‘Jet linear’ was always used as reference, thus leading to 5 runs which were measured in an interleaved way to further reduce bias effects. 20 subjects participated in this experiment.

It is hypothesized that the loudness of the ‘Ratchet wheel sound 2’ and the ‘Spray can’ rattling sound are underestimated by the existing loudness models as in the case of the ratchet wheel sound. This should not be the case for the door creak sound. Figure 9 supports this assumption. For comparison, the results of previous experiments are depicted. The results of the experiments of Parvizian (performed in the same laboratory as the presented experiment) lie in the confidence interval.

Figure 9 also shows results achieved by the hearing model loudness approach but without any consideration of the ACF vs. lag functions. These results are similar to ISO 532-1.

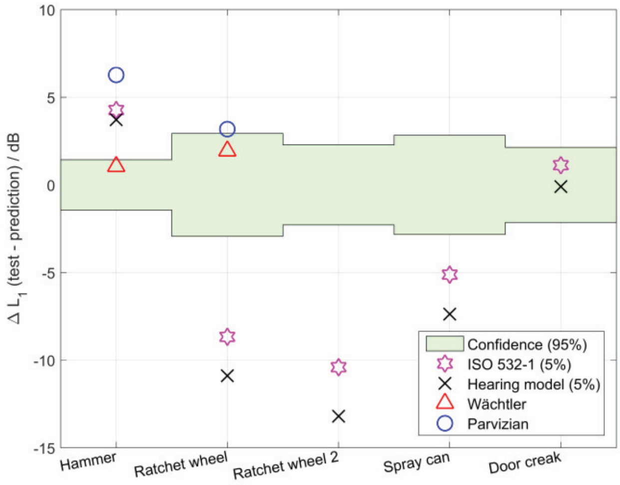

Figure 9: Differences of the results of a loudness matching experiment and the results of ISO 532-1 and the hearing model (as in [16], so far without considering the ACF vs. lag functions) using the indicated signals and a stationary noise as reference. Addionally, the confidence intervals (95 %) of the experiments are depicted. For comparison, the results of previous studies (Wächtler [1], Parvizian [12]) are also given.

6. Conclusions and Outlook

The existing loudness models cannot predict the loudness of all impulsive sounds reliably. For some of the signals, loudness is strongly underestimated (e.g. ratchet wheel sound) while, for other signals, loudness is overestimated (e.g. hammer sound). The proposed new loudness standard ISO 532-1 (based on DIN 45631/A1) performs better than other existing loudness models.

It could be found that a modification of the coarse time structure (e.g. time reversal of single impulses) does not show a strong effect on perceived loudness. Signals with adapted third-octave spectra but different time structures may have significantly different perceived loudness although ISO 532-1 does not predict these differences.

The hearing model analysis of impulsive sounds, especially the evaluation of the three-dimensional autocorrelation analysis (time, lag, and frequency band) permits the derivation of features for an improved loudness prediction of impulsive sounds. Further studies concerning model optimization with initial promising results are ongoing (average deviation from matching results shown in Figure 9 decreases to around 3 dB).

Acknowledgement

The verification of the loudness matching results has been supported by Farzaneh Parvizian during her internship in the research department at HEAD acoustics.

References

[1] M. Wächtler: Wahrnehmung und Modellierung von Lautheit instationärer, technischer Signale, bachelor thesis, Jade Hochschule, Oldenburg, 2012.

[2] J. Rennies, J.L. Verhey, J.E. Appell, B. Kollmeier : Loudness of complex time-varying sounds? A challenge for current loudness models. Proceedings of Meetings on Acoustics, Vol. 19, 050189, 2013.

[3] R. Sottek: Loudness models applied to technical sounds, Noise-Con 2010, Baltimore, 2010.

[4] J. Hots, J. Rennies, J.L. Verhey: Loudness of sounds with a subcritical bandwidth: A challenge to current loudness models? J. Acoust. Soc. Am. 134, EL334–EL339, 2013.

[5] J. Hots, J. Rennies, J.L. Verhey: Loudness of subcritical sounds as a function of bandwidth, center frequency, and level, J. Acoust. Soc. Am. 135 (3), pp. 1313-1320, 2014.

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

[7] J. Rennies, J.L. Verhey, J.L. Chalupper, H. Fastl: Modeling Temporal Effects of Spectral Loudness Summation, Acta Acustica united with Acustica 95, 1112-1122, 2009.

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

[9] DIN 45631/A1:2010. Calculation of loudness level and loudness from the sound spectrum - Zwicker method - Amendment 1: Calculation of the loudness of time-variant sound, Beuth Verlag, 2010.

[10] ISO 532-1. Methods for calculating loudness, Part 1: Zwicker method, in preparation.

[11] ISO 532: 1975. Acoustics – Methods for calculating loudness level, 1975.

[12] F. Parvizian: Internship report, HEAD acoustics GmbH, 2013.

[13] S.D. Ewert: AFC - a modular framework for running psychoacoustic experiments and computational perception models, in Proceedings of the International Conference on Acoustics AIA-DAGA 2013, pp. 1326-1329, Deutsche Gesellschaft für Akustik e.V., Berlin, 2013.

[14] J.L. Verhey: Psychoacoustics of spectro-temporal effects in masking and loudness perception, Phd thesis, University of Oldenburg, Germany, 1999.

[15] J.C.R. Licklider: A Duplex Theory of Pitch Perception, Cellular and Molecular Life Sciences, Vol. 7 (4), pp. 128-134, 1951.

[16] R. Sottek: Modelle zur Signalverarbeitung im menschlichen Gehör. Doctoral Thesis, RWTH Aachen University, 1993.

[17] R. Sottek, F. Kamp, A. Fiebig: A new hearing model approach to tonality, Proc. Internoise, Innsbruck, 2013.

[18] T. Bierbaums, R. Sottek: Modellierung der zeitvarianten Lautheit mit einem Gehörmodel. Deutsche Jahrestagung für Akustik, DAGA 2012, Darmstadt, 2012.

[19] R. Sottek: Progress in calculating tonality of technical sounds, Proc. Internoise, Melbourne, 2014.