



COVER PAGE

Document downloaded by @DAEL

Sat Oct 17 21:45:15 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

A Hearing Model Approach to Time-Varying Loudness, Roland Sottek,
Acta Acustica **vol. 102** (Number 4), 2016, pp. 725-744

DOI

<https://doi.org/10.3813/AAA.918989>

A Hearing Model Approach to Time-Varying Loudness

Roland Sottek

HEAD acoustics GmbH, Ebertstr. 30a, Herzogenrath, 52134, Germany. roland.sottek@head-acoustics.de

Summary

There are different loudness calculation procedures, such as the German standard DIN 45631/A1 and the proposed international standard ISO 532-1 (both based on Zwicker's loudness model) 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 tonal and impulsive sounds do not match the loudness ratings of normal-hearing listeners. To address this issue, the influence of specific signal properties of the sounds on the assessment of loudness was examined focusing on impulsive sounds. On the basis of these experiments, it was studied how far an own improved modeling approach can account for the specific signal properties of these sounds.

PACS no. 43.66.Ba, 43.66.Cb

1. Introduction

Loudness is a very important topic in psychoacoustics because many other psychoacoustic parameters such as sharpness, roughness, and tonality are related to loudness [1]. 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 sounds, whereas models of time-varying loudness are suitable only to a limited extent. For example, recent studies of Rennies *et al.* [2, 3] showed that the predictions of loudness models for tonal sounds and a range of impulsive sounds do not match the ratings of normal-hearing listeners. The challenge of loudness prediction of tonal sounds has already been discussed by Sottek [4] and Hots *et al.* [5, 6].

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 [7], its extension by Rennies *et al.* (DLM ext.) [8], the Time Varying Loudness (TVL) model by Glasberg and Moore using different time constants (short and long, stTVL and ltTVL) [9], and ISO 532-1 [10], replacing ISO 532:1975 section 2 (method B) [11] based on Zwicker's loudness model [12]. The results of the German standard DIN 45631/A1 [13] are not shown because the proposed standard ISO 532-1 corresponds closely with DIN 45631/A1. Thus the results of both standards are very similar. However, ISO 532-1 de-

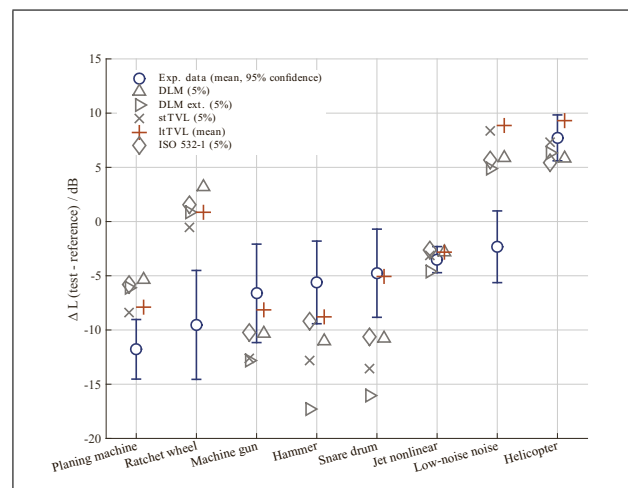


Figure 1. Results of loudness matching experiments using the indicated time-varying sounds and a stationary noise as reference (circles and bars, mean and 95% confidence intervals): the level differences at the point of subjective equality are given. Additionally, the predictions of various loudness models (based on N_5 -values¹ except for ltTVL) are depicted. Data are taken from Rennies *et al.* [2], except the model results of ISO 532-1 have been determined using the audio signals provided by Rennies.

finer the loudness calculation procedure precisely, providing a reference implementation, whereas DIN 45631/A1 is partly imprecise and provides, for the verification of an implementation, only loudness results for a few synthetic tone pulses. Consequently, the reference implementation of ISO 532-1 is preferred over DIN 45631/A1.

Received 11 May 2015,
accepted 6 June 2016.

¹ N_5 : Loudness equalled or exceeded 5% of the time.

None of the employed models could predict all level differences at equal loudness. Especially for the sound *ratchet wheel*, large discrepancies with respect to the predicted level at equal loudness, on the order of 10 dB, apply for all considered models. Moreover, all models underestimate the loudness of the sound *planing machine* by 3 to 6 dB and the loudness of the *low-noise noise* by 7 to 11 dB. They overestimate the loudness of the sound *hammer* by 3 to 12 dB. All models except ItTVL overestimate the loudness of the sound *snare drum* by 6 to 11 dB.

Many loudness calculation procedures can be described with a common scheme (Figure 2). The results vary because of different implementations of each processing block such as the filter bank (number of filters, bandwidth and shape of the filters), the frequency weighting and the nonlinearity between sound pressure and specific loudness.

Loudness calculation begins with the microphone signal $p(t)$. In a preprocessing step the type of sound field, free or diffuse (FF or DF), is considered. This is followed by filtering with a bank of band-pass filters, which in turn is followed by rectifying and low-pass filtering (envelope formation). The next two steps consist of a frequency-dependent weighting g_k and a nonlinear transform from sound pressure or intensity to specific loudness (block N' in Figure 2). For time-varying sounds, the nonlinear decay of the human hearing system must be modeled in detail. Furthermore, the effects of the temporal summation and forward-masking must be taken into account (block NL in Figure 2). Finally, the total loudness $N(t)$ is calculated by summing the specific loudness values. The low-pass filter at the end accounts for the fact that sounds with a duration of 10 ms are perceived as approximately half as loud as sounds with a duration of about 100 ms. Further details of each signal processing block for the proposed standard of time-varying loudness (ISO 532-1) can be found in [14].

The TVL model on the one hand and DLM and ISO 532-1 procedures on the other hand differ mainly in the implementation of the filter bank and in the frequency-dependent weighting g_k .² Whereas the DLM and ISO 532-1 algorithms use 24 filters with a constant bandwidth on the Bark scale (according to Zwicker's loudness model, approximated by 28 3rd-octave filters for ISO 532-1), the TVL algorithm uses approximately 40 filters with a constant bandwidth on the ERB scale. Additionally, the filter shapes are different. The ISO 532-1 algorithm employs very steep 3rd-octave filters of 6th-order and considers the spread of excitation into the neighboring bands by adding to the specific loudness, in each critical band, a slope towards higher frequencies. The resulting specific loudness is the maximum of the specific loudness calculated from the 3rd-octave levels and the value achieved by these slope functions. The TVL model implements rounded exponential filters rather than a kind of "rectangular function".

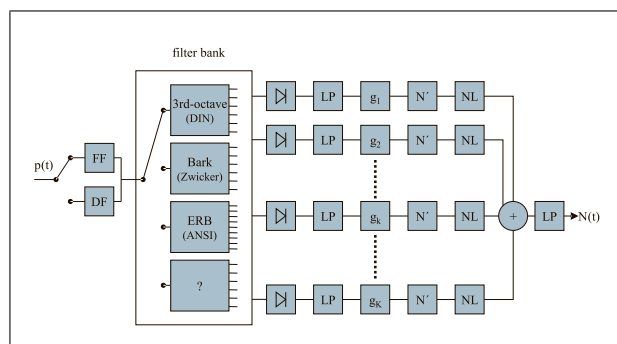


Figure 2. General block diagram of loudness calculation procedures. Block NL (nonlinear network for temporal summation and forward-masking) is used only for the loudness calculation of time varying sounds as in ISO 532-1.

The study of Rennies *et al.* [2] was the motivation to apply an own model to time-varying sounds, especially to the critical impulsive sounds like the sound *ratchet wheel*, and to find an appropriate way to evaluate the performance of the various models. In the following, first, a review of the loudness calculation using the hearing model of Sottek is given, and second, several additional validation experiments are presented. Finally, an advanced approach used for predicting other psychoacoustic parameters (roughness [15], and tonality [16, 17]) is discussed to find features which could be helpful to adjust the hearing model of Sottek for an improved loudness prediction of impulsive sounds.

2. The hearing model (Sottek)

A common way to develop a mathematical model of sound perception is to look at the relations between acoustic stimuli and sensations. Often the result is an empirical formula which can describe (only) a few aspects of sound perception. Because there are so many models for single auditory phenomena, it is very difficult to get an idea of the complete signal processing in human hearing. Thus, it would be desirable to have one principle of overriding importance, which covers many psychoacoustic phenomena. Based on the physiology of the human ear, a theoretical system model was developed with the intention of explaining and describing psychoacoustic effects and basic auditory sensations [15].

It was shown that the developed model could provide a good correlation with the results of many psychoacoustic experiments based on an easy mathematical handling. For example, just-noticeable amplitude and frequency variations were predicted [15]. In this case the nonlinear processing of human hearing plays a fundamental role. Another emphasis of the model lies in the simulation of masking thresholds and in the calculation of basic auditory sensations such as the psychoacoustic parameters loudness, tonality, roughness, and fluctuation strength.

The basic structure of the hearing model is illustrated in Figure 3. The pre-processing consists of filtering with outer and middle ear transfer functions: the

² Here the focus lies on the main differences. For implementation details, please refer to the standards [10, 11, 13] and the journal publications [7, 8, 9, 12].

input signal $p(t)$ is processed by a filter corresponding to the ear's transmission characteristic a_0 of Zwicker's loudness model [1] in series with a high-pass filter of 1st-order (attenuation of 6.5 dB/decade for frequencies below 500 Hz).³ An auditory filter bank consisting of overlapping asymmetric filters⁴ (section 2.1) models the frequency-dependent critical bandwidths and the tuning curves of the frequency-to-place transform of the inner ear, which mediates the firing of the auditory hair cells as the traveling wave from an incoming sound event progresses along the basilar membrane. The inconstant ratio of bandwidth versus frequency of the auditory filter bank conveys a high frequency resolution at low frequencies and a high time resolution at high frequencies, with a very small product of time and frequency resolution at all frequencies, which empowers, for example, human hearing's recognition of short-duration low-frequency events. Subsequent rectification (section 2.2) accounts for the fact that the nerves fire only when the basilar membrane vibrates in a specific direction. The firing rates of the nerve cells are limited to a maximum frequency. This is modeled using 3rd-order low-pass filters (section 2.3). A feature of the hearing model in its usual psychoacoustic application to human perception is a compressive nonlinearity relating sound pressure to perceived loudness. The hearing model spectrum vs. time is accessible as an intermediate result before the implementation of the compressive nonlinearity, e. g., modeled as a power law with piecewise-defined exponents as described in section 2.4, making the algorithm's time/frequency capabilities available as a tool comparable to conventional techniques⁵ regarding level values.

Based on these fundamental processing steps there are a few post-processing mechanisms for the calculation of basic auditory sensations. As one example the loudness results are obtained by summing all specific loudness signals as the output of the compressive nonlinearity (section 2.5).

2.1. Design of the auditory filter bank

2.1.1. Frequency scales

The nonlinear frequency-to-place transform $f = y(z)$ can be deduced directly from the frequency-dependent bandwidth of the auditory filters, assuming that the bandwidth $\Delta f(f)$, independently of the actual frequency, always corresponds to a section of constant length on the linearly divided basilar membrane z -scale. The frequency f results from integration over the bandwidth function $\Delta f(f)$,

³ The attenuation factor is changed from 20 dB/octave (as in [15]) to 6.5 dB/octave for frequencies below 500 Hz in order to better predict the equal loudness contours. Thus, the outer and middle ear filter approximates the inverted equal loudness contour at 100 phon.

⁴ The degree of overlapping depends on the application: e. g., for roughness calculation an overlap of half a critical bandwidth was used, for loudness calculation even no overlap is required, and for a spectral deconvolution algorithm several hundreds of filters were used [15].

⁵ In general, conventional techniques like spectrograms have only a fixed time/frequency resolution.

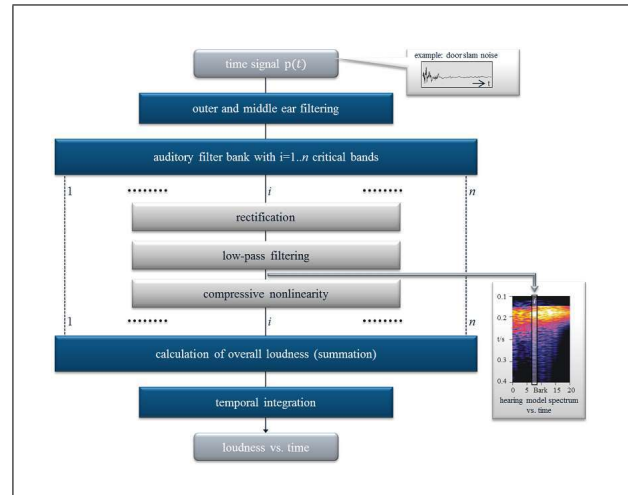


Figure 3. Basic structure of the hearing model for the determination of the hearing model spectrum vs. time and for loudness calculation [15].

which is expressed as a function of the basilar membrane coordinates z :

$$y(z) = \int_0^z \Delta f(y(\tilde{z})) d\tilde{z}. \quad (1)$$

We are looking for $f = y(z)$ or the inverse function $z = y^{-1}(f)$. The solution of equation (1) is given with an arbitrary integration constant C by

$$z = \int \frac{1}{\Delta f(f)} df + C. \quad (2)$$

As an example,

$$\Delta f(f) = \Delta f(f=0) + cf \quad (3)$$

can be used as an approximation to the Bark scale [18] or to the ERB scale [19] (with different constants $\Delta f(f=0)$ and c) [15]. Using this bandwidth function, equation (2) can be rewritten as

$$z = \frac{1}{c} \ln \left(\frac{c}{\Delta f(f=0)} f + 1 \right). \quad (4)$$

Hereby, the constant c is calculated to achieve $y(0) = 0$. The frequency is given by

$$f = \frac{\Delta f(f=0)}{c} (\exp(cz) - 1). \quad (5)$$

It should be noted that according to [20] Greenwood describes the frequency-to-place transform of the basilar membrane based on physiological measurements of Békésy using the following formula:

$$z = 18.31 \log \left(6.046 \frac{f}{\text{kHz}} + 1 \right). \quad (6)$$

The comparison of equation (6) and equation (4) yields for the Greenwood approximation: $c \approx 0.126$ and $\Delta f(f=0) \approx 20.8 \text{ Hz}$.

Table I. Parameters of bandwidth functions shown in Figure 4, optimized in order to minimize the maximum relative error.

$\Delta f(f)$	$\Delta f(f=0)$	c	d/Hz^{-1}
$\Delta f(f=0) + cf$	50.3369	0.1645	
$\sqrt{(\Delta f(f=0))^2 + (cf)^2}$	65.1723	0.1826	
$\Delta f(f=0) + cf + df^2$	77.1853	0.0824	$1.356e^{-5}$

Table II. Parameters of bandwidth functions shown in Figure 5, optimized in order to minimize the mean square of the relative error.

$\Delta f(f)$	$\Delta f(f=0)$	c	d/Hz^{-1}
$\Delta f(f=0) + cf$	50.4764	0.1428	
$\sqrt{(\Delta f(f=0))^2 + (cf)^2}$	81.9289	0.1618	
$\Delta f(f=0) + cf + df^2$	72.9568	0.0844	$1.368e^{-5}$

There are several other analytical bandwidth functions that can be fitted to measured data for which equation (2) can be solved in a closed form [15], e. g.,

$$\Delta f(f) = \sqrt{(\Delta f(f=0))^2 + (cf)^2}, \quad (7)$$

leading to

$$z = \frac{1}{c} \operatorname{arcsinh} \left(\frac{c}{\Delta f(f=0)} f \right) \quad (8)$$

and

$$f = \frac{\Delta f(f=0)}{c} \sinh(cz). \quad (9)$$

Another polynomial bandwidth function is

$$\Delta f(f) = \Delta f(f=0) + cf + df^2, \quad (10)$$

leading to

$$z = \frac{1}{\sqrt{\Delta}} \left(\ln \left(\frac{2df + c - \sqrt{\Delta} c + \sqrt{\Delta}}{2df + c + \sqrt{\Delta} c - \sqrt{\Delta}} \right) \right) \quad (11)$$

with

$$\Delta = c^2 - 4d\Delta f(f=0) > 0! \quad (12)$$

and

$$f = 2\Delta f(f=0) \frac{\exp(\sqrt{\Delta}z) - 1}{c + \sqrt{\Delta} - (c - \sqrt{\Delta}) \exp(\sqrt{\Delta}z)}. \quad (13)$$

Figure 4 shows different approximations to the original measured critical bandwidth values determined by Zwicker *et al.* [21]. The optimization was performed in order to minimize the maximum relative error for the considered frequency range. The maximum relative error using equation (10) is about 8%, a small error considering

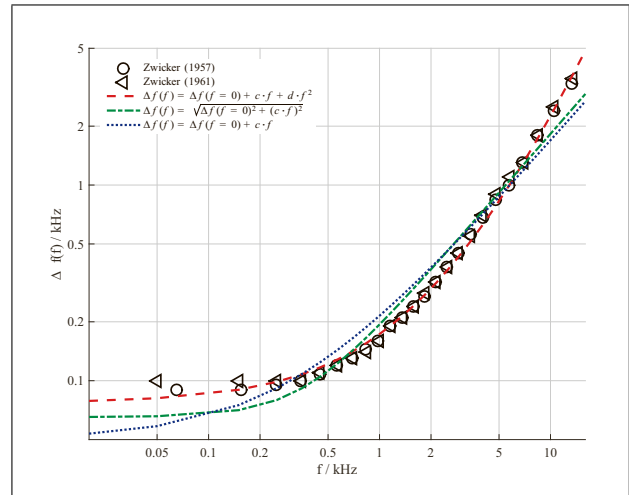


Figure 4. Different approximations for the Bark scale using bandwidth functions according to equations (3), (7) and (10) to the original measured values from Zwicker *et al.* [21], compared to the simplified standardized table [22]. The optimization was performed in order to minimize the maximum relative error for the considered frequency range. The parameters of the bandwidth functions can be found in Table I.

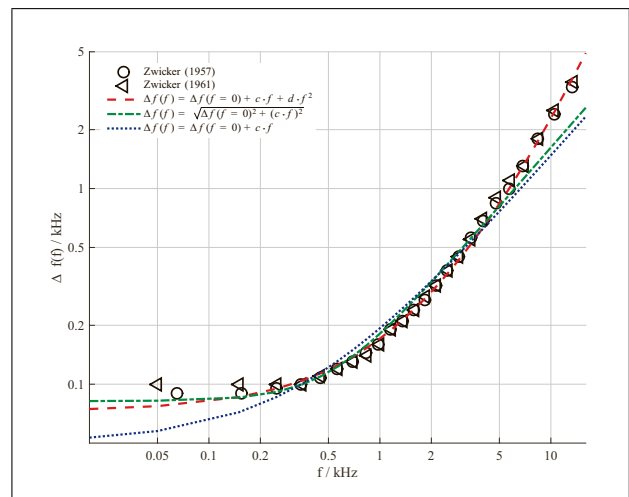


Figure 5. Different approximations for the Bark scale using bandwidth functions according to equations (3), (7) and (10) to the original measured values from Zwicker *et al.* [21], compared to the simplified standardized table [22]. The optimization was performed in order to minimize the mean square of the relative error for the considered frequency range. The parameters of the bandwidth functions can be found in Table II.

that for the tabulated values averaged results from different measurement methods with larger deviations were taken.

Figure 5 shows optimization results minimizing the mean square of the relative error. The approximation for the Bark scale using a bandwidth function according to equation (10) fits very well in both cases to the original measured values from Zwicker *et al.* [21].

A bandwidth function according to equation (10) was chosen first for the ERB-scale with $\Delta f(f=0) = 28.52 \text{ Hz}$, $c = 0.0939$ and $d = 6.23e^{-6} \text{ Hz}^{-1}$; later the

function was changed to a formula according to equation (3).

2.1.2. Shape of the auditory filters

The impulse responses of the auditory filters are chosen as modulated low-pass filters (F is the center frequency of the corresponding critical band),

$$h_F(t) = 2h_{T,F}(t) \exp(j2\pi Ft), \quad (14)$$

using the low-pass function⁶

$$h_{T,F}(t) = \varepsilon(t) \frac{1}{(n-1)!} \frac{1}{\tau(F)} \left(\frac{t}{\tau(F)} \right)^{n-1} \cdot \exp\left(-\frac{t}{\tau(F)}\right), \quad (15)$$

with the corresponding Fourier transform

$$H_{T,F}(f) = \frac{1}{[1 + j2\pi f \tau(F)]^n} = H_T\left(\frac{f}{\Delta f(F)}\right), \quad (16)$$

where n is the filter order⁷ and $\tau(F)$ is a time constant, related to $\Delta f(F)$ by

$$\Delta f(F) = \frac{1}{2^{2n-1}} \binom{2n-2}{n-1} \frac{1}{\tau(F)}. \quad (17)$$

The equivalent rectangular duration Δt_{ERD} is specified as the energy of the normalized impulse response (divided by the maximum $h_{T,F}((n-1)\tau(F))$):

$$\Delta t_{\text{ERD}} = \tau(F) \frac{(2n-2)!}{2^{2n-1}} \left(\frac{e}{n-1} \right)^{2n-2}. \quad (18)$$

The band-pass filtering using $h_F(t)$ as described in equation (14) is implemented in three steps:

1. multiplying the input signal of the filter bank with the corresponding modulator $\exp(-j2\pi Ft)$ (shifting into the low-pass domain, centered at $f = 0$ Hz),
2. filtering the complex signal using $h_{T,F}(t)$, and
3. multiplying the filtered signal with the corresponding modulator $\exp(j2\pi Ft)$ (shifting back to the corresponding frequency F).

Without loss of generality, all hearing model loudness calculations in this paper are performed with the standardized tabulated values of the critical bands according to Zwicker with a resolution of 1 Bark.

2.2. Rectification

In [15] different approaches of rectification were considered: half-wave and full-wave rectification, and finally the calculation of the magnitude of the analytic input signal (by means of the Hilbert transform) with a special feature: it can be shown that the vibration pattern of the basilar membrane may be represented as the convolution of the physical spectrum with a characteristic function, with the

proviso that an adequate frequency transform is used. The neural processing in the brain is considered to be a high-resolution spectral analysis that looks for dominant spectral components and supplies them to a recognition process [15]. In the present study only half-wave rectification is applied.

2.3. Low-pass filtering

The low-pass filtering accounts for the fact that the human ear cannot track envelope fluctuations faster than a certain rate. For the implementation, 3rd-order low-pass filters are used. The cutoff frequency of these filters equals 120 Hz.

2.4. Nonlinear signal processing

The nonlinearity of the auditory system is significant for the loudness perception. The specific loudness distribution, resulting from the application of this nonlinearity to the excitation pattern, also forms the basis for calculating other psychoacoustic parameters such as roughness or fluctuation⁸. Moreover, the nonlinearity mainly determines the just-noticeable amplitude and frequency modulations.

Zwicker [23] determined the level dependence of loudness for 1-kHz tones using the method of adjustment, where test persons had to set the level of a test tone in such a way that it was perceived double or half as loud as a reference tone. The experiment was performed at different levels first for doubling and then for halving loudness.

Zwicker started in a first experiment for half loudness perception at very high levels and then for the next subtests decreased the level of the reference sound subsequently, whereas he did the opposite for double loudness perception. As shown in Figure 6, the mean value of the set-level lies between 10 and 18 dB for a reduction by half, and between 7 and 14 dB for a doubling. At a certain initial level, the scatter range for level reduction is up to 12 dB and at least 4 dB. The higher scatter for level reduction is explained by the influence of a previous tone on the capacity to evaluate subsequent tones.

For simplification, a 10 dB level change for both doubling and halving of loudness was chosen by Zwicker, which corresponds to a power law with an exponent $\nu \approx 0.6$ regarding total loudness related to sound pressure as always in this paper.⁹ The exponent of the power law was set to $\nu = 0.5$ for the calculation of specific loudness values in, e. g., ISO 532-1. Data published by Stevens also show enormous scatter [24].

Due to the considerable scatter it must be doubted whether the participants are at all able to set the loudness (or other perception-related parameters) at certain ratios (for discussion of problems related to loudness scaling refer to [25, 26]).

⁶ $\varepsilon(t)$ is the unit step function.

⁷ Filter order $n = 5$ is recommended.

⁸ Fluctuation and roughness calculations differ only in frequency weighting: a band-pass filter instead of a high-pass filter is applied to the envelope variations.

⁹ For example, a 14 dB level change for both doubling and halving of loudness would correspond to a power law with an exponent of $\nu \approx 0.43$.

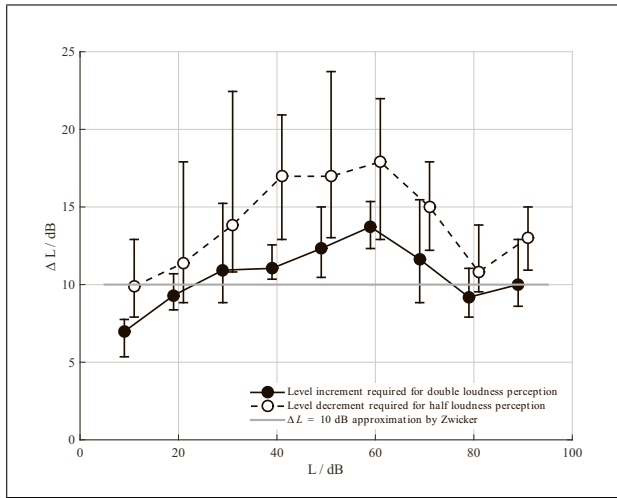


Figure 6. Level increment (filled circles) and decrement (open circles) ΔL of a 1-kHz tone, required for double and half loudness perception, respectively. The diagram shows the median value and the interquartile range of 12 participants (according to Zwicker [23]). Additionally, Zwicker's approximation (horizontal line) is depicted.

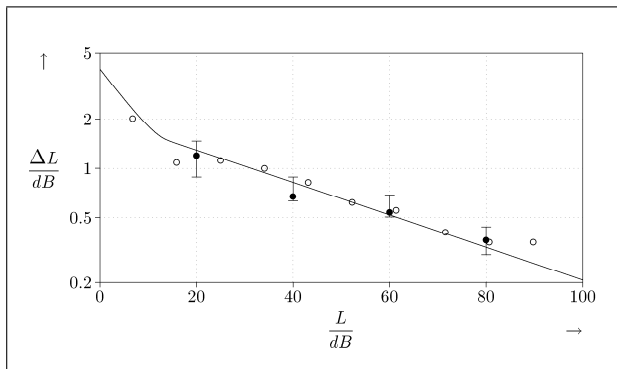


Figure 7. Just-noticeable amplitude differences ΔL for quasi-stationary tone impulses (frequency: 1 kHz). The data taken from Schorer [29] (filled circles: measurements performed by Schorer; open circles: data included from Houtsma *et al.* (1980)) can best be approximated for levels between 15 dB and 80 dB using a power law with the exponent $\nu = 0.2$, replotted from [15].

Garner [27] developed a procedure which constructs the “true” chosen ratio of loudness values (instead of the desired ratio of two) showing finally a much lower exponent $\nu \approx 0.35$. It is also noteworthy that the exponent found by Garner agrees very well with data obtained by measurements of the transverse velocity of the basilar membrane ($0.2 < \nu < 0.3$) [28].

It is much easier for participants to decide whether two sounds generate different or equal sensations, than for them to adjust levels to double the loudness of a reference sound. Following this approach various detection experiments were studied in [15] assuming that the human hearing system analyzes specific loudness in equidistant sensation steps. E.g., presenting two tones with sound pressures $\tilde{p}_1$ and $\tilde{p}_2$, a level change is detected if the maximum specific loudness produced by these tones differs at least by

$\Delta N'$, i.e.

$$\tilde{p}_2 = y^{-1} [y(\tilde{p}_1) + \Delta N'], \quad (19)$$

where y is a nonlinearity function, y^{-1} is the inverse function, $\tilde{p} = p/\sqrt{2}$, and $\Delta N'$ is the equidistant quantization step of the sensation (specific loudness) derived from several data published in the literature. If we plot the level differences using equation (19) and $\tilde{p}_1 = \tilde{p}$

$$\begin{aligned} \frac{\Delta L}{\text{dB}} &= 20 \log(\tilde{p}_2) - 20 \log(\tilde{p}_1) \\ &= 20 \log \{y^{-1} [y(\tilde{p}) + \Delta N']\} \\ &\quad - 20 \log \{y^{-1} [y(\tilde{p})]\} \\ &\approx \frac{20 \Delta N'}{\ln(10)} \frac{d \ln \{y^{-1} [y(\tilde{p})]\}}{dy(\tilde{p})} \\ &= \frac{20 \Delta N'}{\ln(10)} \left(\frac{\tilde{p} dy(\tilde{p})}{d\tilde{p}} \right)^{-1} \end{aligned} \quad (20)$$

in a double logarithmic diagram we get using $y = \tilde{p}^\nu$:

$$\log \left(\frac{\Delta L}{\text{dB}} \right) = \log \left(\frac{20 \Delta N'}{\ln(10) \nu} \right) - \nu \log(\tilde{p}). \quad (21)$$

Figure 7 shows as an example how measured just-noticeable amplitude differences ΔL for quasi-stationary tone impulses can best be approximated by a power law with piecewise-defined exponents ν : $\nu = 1$ for levels near hearing threshold, $\nu \approx 0.2$ at levels between 15 dB and 80 dB (according to equation 21).

In [15] the nonlinear transformation is described by

$$y_1(\tilde{p}) = \begin{cases} \tilde{p}/\tilde{p}_0 - a(\tilde{p}/\tilde{p}_0)^2 & 0 \leq \tilde{p} < \tilde{p}_t, \\ b(\tilde{p}/\tilde{p}_0)^\nu - c & \tilde{p}_t \leq \tilde{p}. \end{cases} \quad (22)$$

The constants a , b and c are determined for the exponent ν in such a way that the function is twice continuously differentiable at the transition $\tilde{p}_t$, and $\tilde{p}_0 = 20 \mu\text{Pa}$.

$$a = \frac{1}{2} \frac{1 - \nu}{2 - \nu} \frac{1}{\tilde{p}_t}, \quad (23)$$

$$b = \frac{1}{\nu} \frac{1}{2 - \nu} \frac{\tilde{p}_t}{\tilde{p}_t^\nu}, \quad (24)$$

$$c = \frac{1}{2} \frac{1 - \nu}{\nu} \tilde{p}_t. \quad (25)$$

Recently, the nonlinearity between specific loudness and sound pressure was reconsidered especially for higher levels above approximately 80 dB according to results of loudness matching experiments using 4th-order band-pass-filtered uniform exciting noise centered around 1 kHz [30]. As described in [30], two noise signals with a bandwidth of 3 Bark and 5 Bark, respectively, were matched in loudness by 24 normal-hearing listeners (8 females/16 males, age between 24 and 50 years). Further experiments with smaller bandwidths were also performed, but disregarded because of the tonality impression of the narrow-band signals (leading to a larger variance in the perceptual data). In a second experiment with 35 normal-hearing

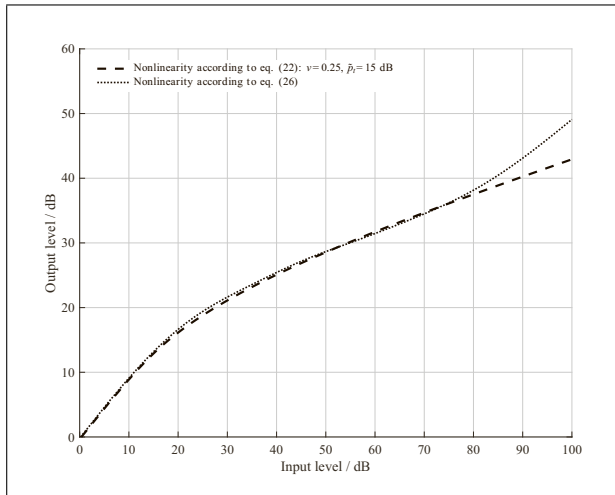


Figure 8. Nonlinearity between specific loudness and sound pressure according to equation (26) (dotted curve) with $\nu_0 = 1$ and 8 exponents ν_i as a result of an optimization procedure. The dashed curve shows the nonlinearity described in equation (22) using only a power law with an exponent $\nu = 0.25$ for levels above 15 dB as in [15].

listeners (13 females/22 males, age between 20 and 35 years), the results of the first experiment were validated. A paired comparison test was performed for each level set, always using the reference sound, the compared sound at the matching level (median value), and at slightly changed levels (adjusted in steps of 1 dB). The points of subjective equality (PSE) found in the first experiment could be confirmed and were slightly corrected [30].

A nonlinearity function y_2 was chosen to simulate this experiment using the presented hearing model [31]

$$y_2(\tilde{p}) = \left(\frac{\tilde{p}}{\tilde{p}_0}\right)^{\nu_0} \prod_{i=1}^N \left(1 + \left(\frac{\tilde{p}}{\tilde{p}_{t_i}}\right)^{\alpha_i}\right)^{(\nu_i - \nu_{i-1})/\alpha_i}, \quad (26)$$

with $\nu_0 = 1$ and N further exponents ν_i . The thresholds $\tilde{p}_{t_i}$ were set to 15 dB, 25 dB, 35 dB, 45 dB, 55 dB, 65 dB, 75 dB, and 85 dB; all α_i were set to 1.5. The $N = 8$ exponents ($\nu_1 = 0.6602$, $\nu_2 = 0.0864$, $\nu_3 = 0.6384$, $\nu_4 = 0.0328$, $\nu_5 = 0.4068$, $\nu_6 = 0.2082$, $\nu_7 = 0.3994$ and $\nu_8 = 0.6434$) were achieved by applying an optimization procedure in order to minimize the mean squared error between the results of the loudness matching experiment and the results of the model calculation.

Figure 8 compares the previous and extended nonlinearities. The steeper slope at higher levels agrees with results of experiments from Buus *et al.* [32] and Epstein *et al.* [33].

There is more evidence for a more compressive nonlinear behavior of the human hearing system (power law with an exponent $\nu \approx 0.2 \dots 0.25$) than the one used in common loudness models, coming from the detection of just-noticeable amplitude modulation (see section 2.6), frequency modulation [15] and the level dependence of roughness calculation (based on the hearing model) [15]. In this case the nonlinear processing is modeled using an

even more compressive nonlinearity applied to the result of a correlation analysis, which for stationary signals behaves as the above-described power law (see next section); for details to roughness calculation refer to [34, 35].

As a consequence of using another nonlinear relation between specific loudness and sound pressure than in current standards, the output values differ, e. g., from the results of ISO 532-1. A level increment of 10 dB will no longer cause the doubling of the loudness of 1 kHz tones, often considered as a test case for loudness models based on an oversimplifying assumption.

2.5. Temporal integration of loudness

Different approaches of temporal integration were implemented and studied. In general, 3rd-order low-pass filters provide a good fit to loudness data for sounds with varying length (section 2.7). The impulse responses of 3rd-order low-pass filters have a similar shape as the ear's temporal window proposed by Moore *et al.* [36] (weighted sum of two rounded exponentials, asymmetrical shape using different time constants for the window parts to the left and the right of the maximum, respectively), but the implementation is more practical [15]. Vogel also proposed a 3rd-order low-pass filter for temporal integration of loudness [37].

Amplitude- as well as frequency-modulated tones generate envelope fluctuations in the individual channels of the auditory filter bank. These fluctuations are analyzed with respect to maxima and minima of the specific loudness values. It is assumed that the fluctuations can be detected if the difference between the extremes is at least $\Delta N'$.

For simulating the influence of modulation frequency on the just-noticeable amplitude and frequency modulations, the time weighting ("the ear's temporal window") is of particular importance. Figure 9 compares results using the described hearing model with published data from measurements (by Schorer [29] and Riesz (1928, included in [29]) as well as by Maiwald [38]). The simulations were performed using a 3rd-order low-pass filter according to equation (15) for temporal integration [15] and as nonlinearity a power law with the exponent $\nu = 0.25$ for medium levels.

2.6. Autocorrelation analysis

Inspired by the work of Licklider [39] concerning human pitch perception, the sliding autocorrelation function was used as a processing block in the hearing model of Sottek for the calculation of roughness and fluctuation strength [15] and later also for other psychoacoustic quantities like tonality [16, 17]. Figure 10 displays the basic hearing model structure and Figure 11 an example of the sliding autocorrelation function (ACF) as described by

$$\hat{\phi}_{ss}(t, \tau, i) = \int_{-\infty}^{\infty} s_i(\tilde{t}) \cdot s_i(\tilde{t} - \tau) \cdot \tilde{w}_i(\tilde{t} - \tau) d\tilde{t}, \quad (27)$$

where τ is the lag, and s_i is the band-pass signal of the i th critical band. The weighting function $\tilde{w}_i(t)$ determines

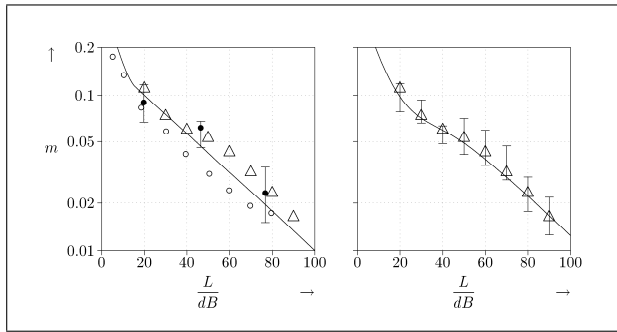


Figure 9. Just-noticeable amplitude modulation at $f = 1$ kHz and sinusoidal modulation with $f_{\text{mod}} = 4$ Hz, data taken from Schorer [29] (filled circles: median value and interquartile range of measurements performed by Schorer; open circles: data included from Riesz (1928)) as well as by Maiwald [38] (triangles, median value) compared with simulation results using as nonlinearity a power law with the exponent $\nu = 0.25$ for levels above 15 dB (left), replotted from [15]. Additionally, the data taken from Maiwald [38] are fitted even better with a slightly modified nonlinearity (right, median value and interquartile range), replotted from [15].

how the past signal values contribute to the actual value of the autocorrelation function. Defining $w_i(t) = \tilde{w}_i(-t)$, equation (27) can be expressed advantageously as a convolution product:

$$\hat{\phi}_{ss}(t, \tau, i) = [s_i(t) \cdot s_i(t - \tau)] * w_i(t). \quad (28)$$

For $w_i(t)$ an impulse response of a 3rd-order low-pass filter is used.

The calculation of the ACF according to equation (28) is indeed highly efficient compared to the integral formulation in equation (27), but still time consuming. Therefore, the sliding ACF is calculated block-wise using the DFT to shorten computing time. In this case, the weighting of the past signal values is a little different. First, there is a low-pass effect due to averaging over the block length. Second, additional 3rd-order low-pass filters are applied to the ACF time curve for each frequency band and lag of interest.¹⁰

In the first hearing model loudness approach [15], the calculation was based only on a summation of compressed band-pass signals (Figure 3). Recently, it was proposed to use the autocorrelation function of the band-pass 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 [17]. Without weighting the loudness of tonal components differently, the specific loudness is calculated by applying the compressive nonlinearity to the square root of $\text{ACF}(\tau = 0)$, because the nonlinear transform is defined in this paper for pressure input signals, whereas $\text{ACF}(\tau = 0)$ corresponds to the energy of the input signal.

¹⁰ For all simulations in the following a sampling rate of 48 kHz and a DFT length of 1024 is chosen. An overlap of approximately 75% is used. The equivalent rectangular duration of the 3rd-order low-pass filters amounts to 5 ms.

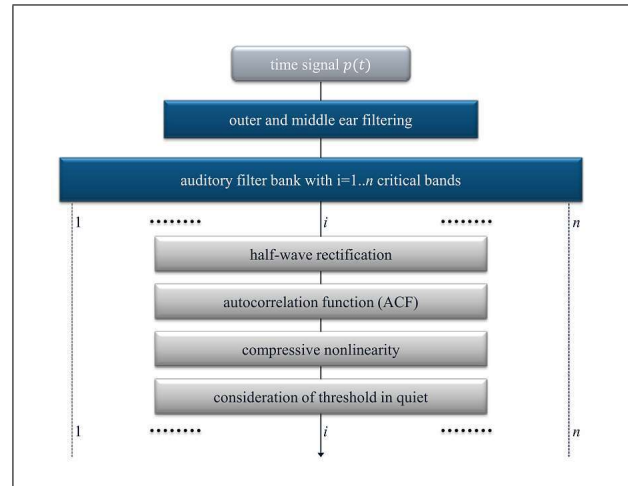


Figure 10. Basic hearing model structure including the determination of the autocorrelation function as function of time, lag and frequency band (see also Figure 11 and compare to Figure 3).

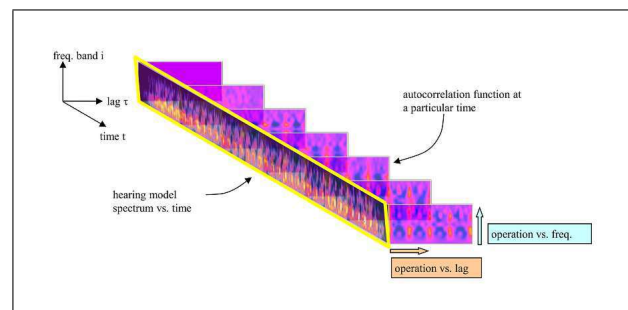


Figure 11. Sliding autocorrelation function applied to the band-pass signals. The hearing model spectrum vs. time is accessible as $\text{ACF}(\tau = 0)$ as a function of time and frequency band. Different operations can be applied vs. lag and vs. frequency bands for further processing.

The specific loudness in each band is reduced by the specific loudness values corresponding to the lower threshold of hearing for pure tones minus a fixed level value (for the first attempt 6 dB) and set to zero for negative results. Thus a pure tone at the threshold would cause a nonzero specific loudness. The fact that the tone cannot be detected could be taken into account by an additional threshold for overall loudness. This method makes it possible to model the decreased threshold for complex sounds like multiple tones of different frequency (each tone at threshold). This simple but effective approach must be studied and perhaps optimized later together with the different weighting of the loudness of tonal and non-tonal components.

Afterwards, the entire ACF is normalized for each lag with the same factor, i.e., the ratio of the specific loudness and $\text{ACF}(\tau = 0)$.

The three-dimensional autocorrelation 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.

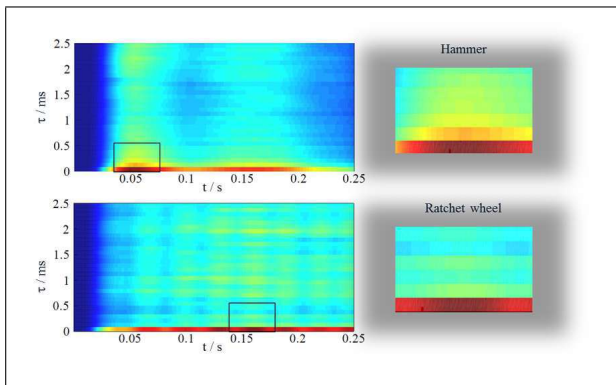


Figure 12. Sum of the autocorrelation functions of all critical bands as a function of time t and lag τ for the sound *hammer* (upper) and the sound *ratchet wheel* (lower) used in the Rennies *et al.* study [2]. The zoomed areas at the right side better show the different decays of the ACF vs. τ for the two sounds.

Figure 12 shows that the decay of the ACF is much steeper for the sound *ratchet wheel* than for the sound *hammer* (sounds used in the Rennies *et al.* study [2], see Figure 1). Additionally, the maximum of the normalized ACF in a certain lag range ($\tau > 0$) is lower for the sound *ratchet wheel*. Parameters related only to specific frequency ranges (low, middle, and high frequencies) could be used as further features for a recognition and an improved loudness evaluation of such ‘critical’ impulsive sounds.

2.7. Model adjustment with synthetic sounds

In the following, several published experiments using synthetic sounds were studied in order to validate or adjust the parameters of the nonlinearity described in equation (26) and the parameters of the low-pass filter for temporal integration.

In section 2.4 a first experiment of Zwicker related to the level dependence of loudness for 1-kHz tones is described, showing differences between the results for halving and doubling loudness. Zwicker reviewed the design of his experiment and performed another experiment where he started for half and double loudness perception always with the lowest level and then increased the reference level by 10 dB for the next sub-test.

These results, depicted in Figure 13, show comparable scatter ranges for half and double loudness perception, respectively. In this case, the mean value of the set-level lies between 5 and 13 dB for a reduction by half. The experimental results are predicted very well (rms error: 0.9 dB, Figure 13) by the hearing model of Sottek using an optimized nonlinearity according to equation (26) (dashed-dot curve in Figure 14) with 9 different exponents ($N = 9$). The thresholds $\tilde{p}_{i_l}$ were set to 10 dB, 20 dB, 30 dB, 40 dB, 50 dB, 60 dB, 70 dB, 80 dB and 90 dB; all α_i were set to 1.5, and the exponents to $\nu_1 = 0.5303$, $\nu_2 = 0.4518$, $\nu_3 = 0.3723$, $\nu_4 = 0.6024$, $\nu_5 = 0.1132$, $\nu_6 = 0.1274$, $\nu_7 = 0.8258$, $\nu_8 = 1.0034$ and $\nu_9 = 0.4227$. Nevertheless, also the simulation results using the nonlinearity ac-

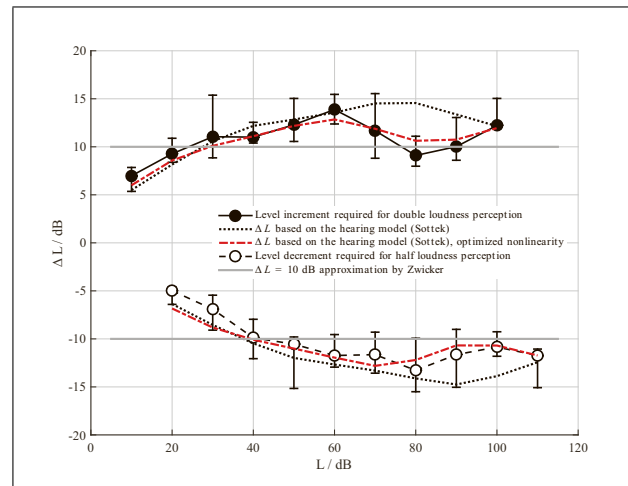


Figure 13. Level increment (filled circles) and decrement (open circles) ΔL of a 1-kHz tone, required for double and half loudness perception, respectively, always starting with the lowest level. The diagram shows the median value and the interquartile range of 12 participants (according to Zwicker [23]). Additionally, Zwicker’s approximation (horizontal lines) is depicted, and as well as the results of simulations based on the hearing model of Sottek using two different nonlinearities according to equation (26) (see text below).

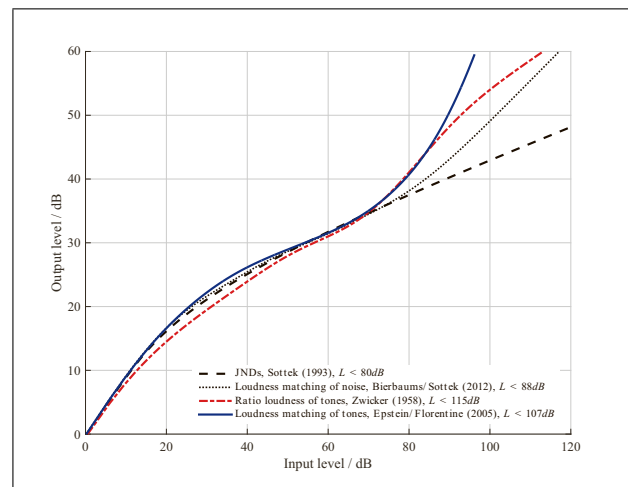


Figure 14. Nonlinearities between specific loudness and sound pressure derived from different data: JNDs [15], loudness matching of noise [30], ratio loudness of tones [23], and from loudness matching of tones [33]. The dashed curve shows a nonlinearity that can be described by a linear function for levels below about 15 dB and a power law with an exponent $\nu = 0.25$ for higher levels (equation 22). The other three nonlinearities are functions according to equation (26); they are results of optimization procedures to fit specific experimental data with the lowest mean squared error. Additionally, the maximum level used in the experiments is indicated.

ording to equation (26) derived from loudness matching results of noise signals (dotted curve in Figure 14, maximum level used: 88 dB) agree well with the experimental data (rms error: 2.1 dB, Figure 13).

The solid curve in Figure 14 shows another nonlinearity according to equation (26) optimized for loudness matching experiments using 1-kHz tones of two different durations (5 ms and 200 ms) performed by Buus *et al.* [32] as well as by Epstein and Florentine [33]. For such experiments using very short impulses, the simulation results depend on the nonlinearity as well as on the temporal integration. An overview about studies showing that the loudness function is less steep at moderate levels than at low or high levels is given by Florentine and Epstein [40].

For the temporal integration, other approaches besides 3rd-order low-pass filters are considered, e. g., the use of two parallel 1st-order low-pass filters with different weighting factors and time constants, for example in ISO 532-1 as described in [14] and by Hots *et al.* [41]. The study by Hots *et al.* compared different models using up to two 1st-order filters in a serial or parallel arrangement (no 3rd-order low-pass filters). The advantage of a parallel arrangement of two filters was shown for loudness matching data from an international round robin test¹¹ using one data set (1-kHz tones of different durations for a level of 55 dB) applying an extended DLM model [8].

The performance of the different approaches is shown in the following using a nonlinearity according to equation (26) (solid curve in Figure 14) with 4 different exponents ($N = 4$). The thresholds $\tilde{p}_i$ were set to 15 dB, 35 dB, 70 dB, and 90 dB; α_i was set to 1.5, and the exponents to $\nu_1 = 0.5366$, $\nu_2 = 0.2188$, $\nu_3 = 0.5638$ and $\nu_4 = 2$.

The simulation results shown in Figure 15 to Figure 18 are based on the assumption that at the point of equal loudness the maxima of the loudness vs. time functions of test and reference sounds are identical. Figure 15 and Figure 16 show results of loudness matching experiments of 1-kHz tones of different durations from Buus *et al.* [32] and hearing model calculations using several values for the equivalent rectangular duration Δt_{ERD} of the 3rd-order low-pass filter for temporal loudness integration as defined in equation (18). The simulation results using $\Delta t_{ERD} = 60$ ms agree very well with the experimental results of Buus *et al.* [32] as well as with the loudness matching results of Epstein and Florentine [33] shown in Figure 17.

In the loudness matching experiments performed by Buus *et al.* [32] and by Epstein and Florentine [33], very short and long tone pulses were adjusted in loudness whereas in the experiments performed by Poulsen [42] only pairs of double length were compared. That means the nonlinearity does not have such a strong influence on the simulation results of Poulsen's experiment. Often, accumulated data are shown, based on pairwise comparisons (5 ms/10 ms, 10 ms/20 ms, etc.) as, e. g., in [42]. This could be misleading as it suggests that the experiment was performed using a different setup with respect to durations and levels. To avoid confusion, Figure 18 shows the measured experimental data derived from Poulsen [42] as dif-

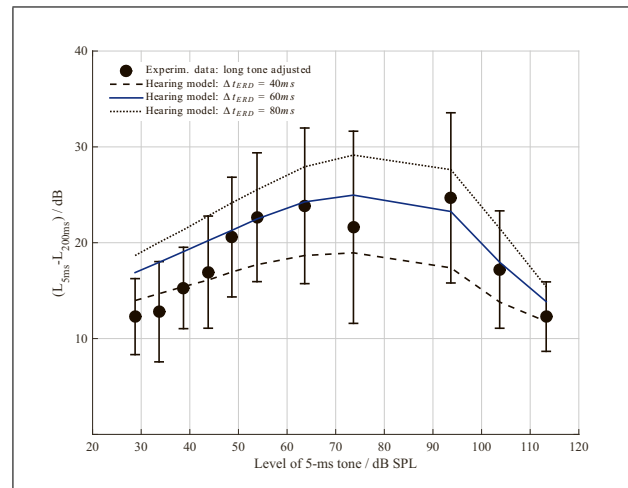


Figure 15. Results of loudness matching of 1-kHz tones of different durations (5 ms and 200 ms, long tone adjusted) from Buus *et al.* [32] (averaged across six participants) and hearing model calculations with different values for Δt_{ERD} of the 3rd-order low-pass filter for temporal loudness integration: 40 ms, 60 ms, and 80 ms.

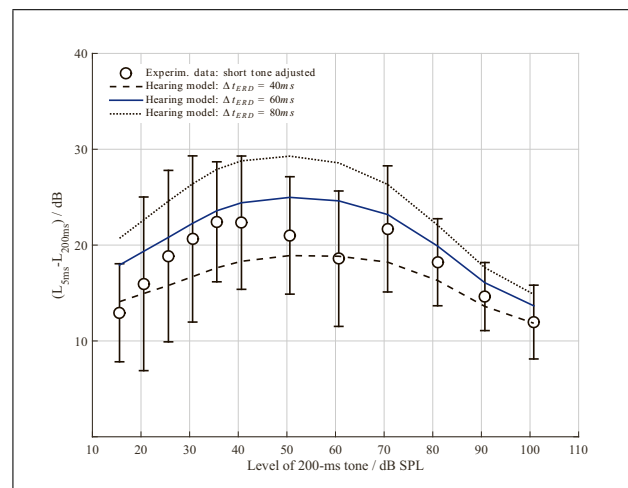


Figure 16. Results of loudness matching of 1-kHz tones of different durations (5 ms and 200 ms, short tone adjusted) from Buus *et al.* [32] (averaged across six participants) and hearing model calculations with different values for Δt_{ERD} of the 3rd-order low-pass filter for temporal loudness integration: 40 ms, 60 ms, and 80 ms.

ferences between two neighboring data points in Poulsen's data set. The simulation results are based on the comparisons in the original experiment (for each comparison the level of the longer stimulus was set to the reference level of interest: 35 dB). Even for longer durations (160/320 ms and 320/640 ms), the experimental data (Figure 18) show a mean level difference of about 1.5 dB for the shorter tone at the point of subjective equality, but the 95% confidence interval starts around 0 dB.

The 3rd-order low-pass filters are in general well suited in combination with an appropriate nonlinearity to describe the temporal loudness integration. An equivalent rectangular duration of $\Delta \tau_{ERD} = 60$ ms is used to calculate

¹¹ Data were collected from a very large group of subjects in a lot of laboratories, each using one of three different measurement procedures.

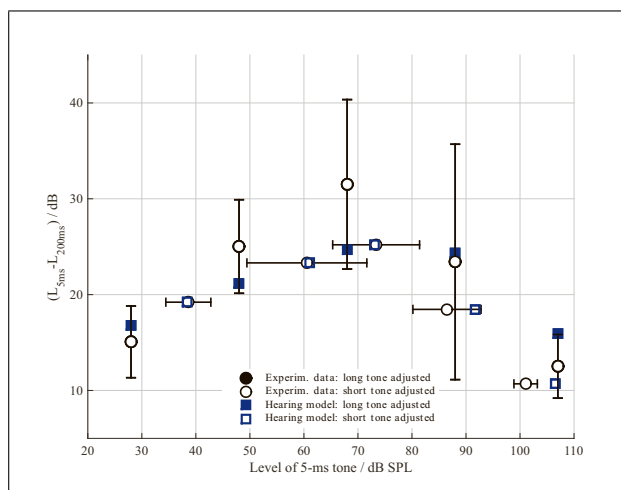


Figure 17. Results of loudness matching of 1-kHz tones of different durations (5 ms and 200 ms, short tone as well as long tone adjusted) from Epstein and Florentine [33] (mean and standard deviation of nine participants) and hearing model simulations. The equivalent rectangular duration of the 3rd-order low-pass filter for temporal loudness integration equals $\Delta t_{\text{ERD}} = 60$ ms.

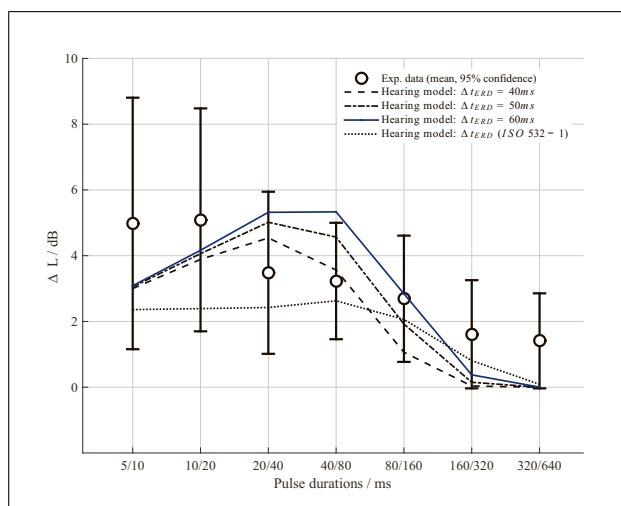


Figure 18. Results of loudness matching of 1-kHz tones of different durations from Poulsen [42] for a level of 35 dB together with hearing model calculations using a 3rd-order low-pass filter for temporal loudness integration with different values of Δt_{ERD} (40 ms, 50 ms, and 60 ms) and two parallel 1st-order low-pass filters as proposed for ISO 532-1 [14].

the loudness of the technical sounds in the next section. The data provided by Poulsen can be better approximated by a slightly shorter equivalent rectangular duration. Prediction results using a temporal integration as defined in ISO 532-1 based on two parallel 1st-order low-pass filters (time constants 3.5 ms and 70 ms)¹² lie also within the confidence interval, but are deviating more for shorter durations (5 ms/10 ms, 10 ms/20 ms) from the mean values of the experiment (Figure 18).

¹² The total loudness is the weighted sum (factors 0.47 and 0.53) of the filtered signals.

There is a need for further research on temporal integration: for example it could be verified whether a level-dependent time constant (longer time constants for higher levels) could better fit the experimental data.

3. Validation of the hearing model using technical sounds

The hearing model approach to time varying loudness as described in section 2 shall be validated with results from four loudness matching experiments using an adaptive two-alternative forced-choice procedure applied to technical sounds, especially focusing on impulsive sounds, and further, the effectiveness of the hearing model loudness calculation shall be compared to the performance of other loudness models. The hearing model parameters are adjusted based on several fundamental experiments using synthetic sounds as illustrated in section 2.7.¹³ The design of all four experiments is almost the same, except for Experiment 1, which was performed by other authors (Rennies *et al.* [3], using diffuse-field instead of free-field equalized headphones as in Experiments 2 to 4). Experiment 2 is a replication of Experiment 1. For Experiments 3 and 4 new stimuli were used. Therefore, a common description of the procedure is given in the following.

3.1. Stimuli

Table III lists all sounds and their sources. Check marks indicate in which of the four experiments they were used. Sounds 1 to 10 were provided by Rennies together with the experimental data of Experiment 1 [3], some other sounds were taken from a free online sound data base and sound *white noise* (WN) was generated using the software HEAD acoustics ArtemiS. Sound durations were between 1.6 s and 2.62 s.

Experiment 1 was performed by Rennies *et al.* [3]. The technical signals used in Experiment 1 ranged from rather stationary to highly impulsive sounds.

Experiment 2 is a replication of Experiment 1 using free-field equalized headphones.

In Experiment 3 the three sounds *ratchet wheel*, *hammer* and *white noise* (WN) were used, all adjusted to the overall sound pressure level of 47.7 dB. These sounds were filtered to match the 3rd-octave spectra of the two other sounds, thus leading to three groups of three sounds, e.g., *ratchet wheel* (original), *white noise* (filtered, 3rd-octave spectrum of the sound *ratchet wheel*), *hammer* (filtered, 3rd-octave spectrum of the sound *ratchet wheel*) etc. The stimuli in each of the three groups approximately preserved the temporal structure of the original sounds, but had a highly similar long-term 3rd-octave spectrum (namely that of the original sound in this group).

Experiment 4 is similar to Experiments 1 and 2 but with new time-varying signals. The motivation for this experiment was to find sounds with similar characteristics as the

¹³ A short overview of the model parameters used for the predictions is given in section 3.4.

Table III. List of all sounds and their sources used in the different experiments. freesound: see <https://freesound.org>

No.	Name	Exp. 1	Exp. 2	Exp. 3	Exp. 4	Reference
1	jet linear (reference)	✓	✓		✓	Gee <i>et al.</i> (2007)
2	planing machine	✓	✓			Brennecke and Remmers (1983)
3	ratchet wheel	✓	✓	✓	✓	freesound: atomata, 26869, 0–2.02 s
4	machine gun	✓	✓			freesound: Matt_G, 30749
5	hammer	✓	✓	✓	✓	freesound: cognito perceptu, 17012
6	snare drum	✓	✓			Rennies <i>et al.</i>
7	impact wrench	✓				Brennecke and Remmers (1983)
8	jet nonlinear	✓	✓			Gee <i>et al.</i> (2007)
9	screw wrench	✓				freesound: digifishmusic, 54264
10	helicopter	✓	✓			freesound: klankbeeld, 116692
11	white noise (WN)			✓		HEAD acoustics ArtemiS
12	ratchet wheel 2				✓	freesound: atomata, 26869, 2.5–4.52 s
13	spray can				✓	freesound: willc2_45220, 73266
14	door creak				✓	freesound: plagasul, 386

sound *ratchet wheel* in order to investigate if other features of the ACF for the hearing model loudness of Sottek could be used to improve the loudness calculation as mentioned in section 2.6.

3.2. Procedure

Loudness matching was performed in all experiments using an adaptive two-alternative forced-choice procedure (for Experiments 2 to 4 the AFC-toolbox provided by Ewert [43] was used). The participants heard a reference sound (sound pressure level: 60.8 dB) and a test sound separated by 500 ms of silence, presented diotically via a digital equalizer (PEQ V from HEAD acoustics with free-field equalization) and Sennheiser HD 650 headphones¹⁴ in a sound-attenuated listening studio. The order of test and reference sound was randomly changed. The participants were asked to indicate which sound was perceived as louder. The sound pressure level of the test sound was adjusted depending on the participant's response. The initial step size was 8 dB. The step size was halved after each upper reversal until a step size of 2 dB was reached. The matching was stopped when the results had converged to almost the same perceived loudness of the test and the reference sound, i. e., after four further reversals. The mean value of all levels after reaching the minimum step size was taken as point of subjective equality. Experiment 2 was performed three times for each test sound using different starting values in order to reduce bias effects. For this experiment all the runs were divided into sessions of maximum 6 runs each with different sounds, which were measured in an interleaved way to further reduce bias effects [44]. The different starting points led to almost the same results (see below). In Experiments 3 and 4, performed by the same participants, only one starting level was used.

For all experiments, the same reference sound *jet linear* was used with the test sounds indicated in Table III, except for Experiment 3 where three reference sounds (*ratchet*

wheel, *hammer*, and *white noise*) were used. The corresponding test sounds for each reference sound were the other two reference sounds, but filtered to have the same 3rd-octave spectra as the reference sound as described in section 3.1.

3.3. Participants

Eleven normal-hearing listeners participated in Experiment 1. None of these participants reported any hearing difficulties [3]. Fifteen normal-hearing listeners with self-reported normal hearing (including the author) participated in the other experiments, all in Experiment 2, nine of them in Experiment 3 and eleven in Experiment 4.

3.4. Loudness models

For each model, predictions were performed similar to the experiment, adjusting the level of the test sound in order to match the loudness of the reference sound. Silent intervals of 100 ms were added at the beginning and end of the sound, only for processing but were not considered for the representative value of overall loudness. This was also done in [3].

For Experiment 1, the model predictions were performed using the sounds described in Table III pre-processed by a filter provided by Rennies in order to mimic the diffuse-field equalization of the HD 650 headphones used in this experiment. This was not necessary for the other own experiments using a free-field equalizer for playback.

Four different time-varying loudness models were used to predict the experimental data.

3.4.1. ISO 532-1 (DIN 45631/A1)

The general structure of ISO 532-1 and DIN 45631/A1 are described in the introduction. The algorithm considers spectral loudness summation and forward masking. Temporal integration is performed using the weighted sum of two parallel low-pass filters with different time constants [14]. The loudness value equalled or exceeded 5% of the time is taken as representative value for overall loudness.

¹⁴ In Experiment 1 performed by Rennies *et al.* [3] no additional equalizer was used.

3.4.2. DLM

The Dynamic Loudness Model as proposed in [7] has a structure comparable to ISO 532-1 and DIN 45631/A1, considering spectral loudness summation in short intervals and subsequent low-pass filtering (8 Hz). The loudness value equalled or exceeded 5% of the time is taken as representative value for overall loudness.

3.4.3. TVL (stTVL, mtTVL, ltTVL)

The general structure of TVL is similar to the previously described methods to calculate the instantaneous loudness, which is integrated using a kind of automatic gain control using a short attack time constant (22 ms) and a longer release time constant (50 ms) to calculate the short-term loudness stTVL. There is another version of TVL, long-term loudness ltTVL, using a second temporal smoothing stage with an attack time constant of 99 ms and a release time constant of 2 s, applied to the short-term loudness. Only stTVL and ltTVL are mentioned in the original paper [9], but the software, available for download¹⁵ until 16th September 2015, was using a release time constant of 200 ms for ltTVL, although there was a notice stating that this version corresponds to [9]. In the meantime this has been corrected. Nevertheless, the ‘faulty’ TVL version showed good prediction results (section 3.5). Therefore, this TVL model has been included as mtTVL, providing further insights regarding the release time constant of the TVL model. The loudness value equalled or exceeded 5% of the time is taken as representative value for overall loudness in the case of stTVL, mtTVL and ltTVL; additionally, the mean value of mtTVL and ltTVL: five values, whereas normally only two are considered: stTVL (5%) and ltTVL (mean).

3.4.4. HML

The hearing model loudness of Sottek is described in section 2. The model parameters used for the predictions are (see Figure 10)¹⁶:

- Outer and middle ear filtering corresponding to the ear’s transmission characteristic α_0 of Zwicker’s loudness model [1] in series with a high-pass filter of 1st-order (attenuation of 6.5 dB/decade for frequencies below 500 Hz),
- Auditory filter bank consisting of 24 non-overlapping critical band filters as described in section 2.1 (filter order $n = 5$, tabulated values of center frequency and bandwidth as in [1]),
- Half-wave rectification,
- Sliding autocorrelation function, block-wise calculation by means of a DFT (length: 1024, overlap: 75%), 3rd-order low-pass filters ($\Delta t_{\text{ERD}} = 5$ ms) applied to the ACF vs. time curve for each frequency band and lag of interest,

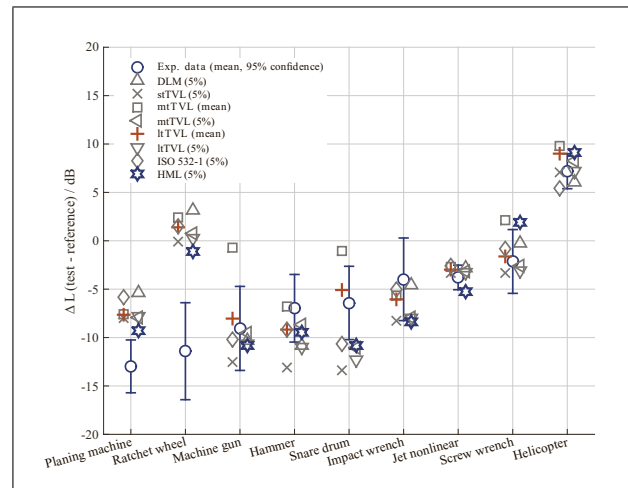


Figure 19. Results of loudness matching experiments using the indicated time-varying sounds and a stationary noise as reference (circles and bars, mean and 95% confidence intervals); the level differences at the point of subjective equality are given. Additionally, the predictions of various loudness models (based on N_5 - or mean values) are depicted.

- Nonlinearity according to equation (26) derived from loudness matching results of noise signals (dotted curve in Figure 14),
- Subtraction of the specific loudness value corresponding to the lower threshold of hearing for pure tones reduced by 6 dB from the calculated specific loudness in each band (negative values are set to zero),
- Temporal integration using 3rd-order low-pass filters with an equivalent rectangular duration of $\Delta t_{\text{ERD}} = 60$ ms.

The loudness value equalled or exceeded 5% of the time is taken as representative value for overall loudness.

3.5. Results

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

The open circles in each panel of Figure 19 show mean level differences between the test sounds and the reference sound at equal loudness for the indicated sounds, based on averaged values for each participant and each test sound using three different starting values. Additionally 95% confidence intervals are given. The data were provided by Rennie [3]. The sounds are sorted according to their level difference at equal loudness. The predictions of the different models described above are also depicted in the same Figure 19. The results of Rennie *et al.* [3] could be verified for DLM (5%), stTVL (5%) and ltTVL (mean). The other models were not used in [3].

The largest deviations were observed for the sounds *ratchet wheel* and *planing machine* as in a previous experiment of Rennie *et al.* [2] (Figure 1). All models underestimate the loudness of these sounds. Almost all models (except mtTVL (mean)) overestimate the loudness of the sound *hammer*, and the loudness of the sound *snare drum* is overestimated by almost all models (except mtTVL

¹⁵ <http://hearing.psychol.cam.ac.uk/demos/demos.html>

¹⁶ All time signals are resampled to 48 kHz.

Table IV. Coefficient of determination (R^2) and Spearman's rank correlation coefficient (ρ_s) for Experiment 1 using the rms level differences ΔL and A-weighted level differences $\Delta L(A)$.

	mtTVL (mean)	DLM (5%)	stTVL (5%)	ISO 532-1 (5%)	ltTVL (5%)	mtTVL (5%)	ltTVL (mean)	HML (5%)
$R^2(\Delta L)$	0.45	0.26	0.36	0.31	0.38	0.42	0.48	0.53
$R^2(\Delta L(A))$	0.17	0.01	0.01	0.01	0.02	0.03	0.11	0.22
$\rho_s(\Delta L)$	0.43	0.43	0.35	0.45	0.35	0.37	0.53	0.58
$\rho_s(\Delta L(A))$	0.43	0.18	0.05	0.22	0.03	0.07	0.40	0.47

(mean) and ltTVL (mean)). In general, the deviations between experimental results and predictions were larger in the previous experiment.

To assess the overall prediction quality of the models the coefficient of determination (R^2), Spearman's rank correlation coefficient (ρ_s), the root mean squared error (ϵ), and the prediction bias (mean of prediction error) (θ) are proposed in [3].

Which of these quantities is appropriate for predicting the quality of loudness models with respect to loudness matching experiments? Obviously, a very low root mean squared error is a good indicator: preferably ϵ should be close to zero or the prediction should lie within the 95% confidence interval because the exact mean of the experimental data is not known.¹⁷ In addition, a small prediction bias θ is a meaningful indicator.

In loudness matching experiments the participants are normally asked to indicate which sound is perceived as louder, e. g., the reference or the test sound. Subsequently, the test sound is multiplied by a factor depending on the participant's response. This action applies to the unweighted and any weighted level such as the A-weighted sound pressure level of test and reference signals in the same way. Therefore, the choice of the parameter to be evaluated appears arbitrary. Typically, the unweighted level is considered as in the paper of Rennie *et al.* [3], but often without justification.

Table IV shows the correlation-based parameters R^2 and ρ_s for Experiment 1 using the rms level differences ΔL and also A-weighted level differences ($\Delta L(A)$), by adding to the result of each test sound a different constant: the difference of the A-weighted level of the reference and the test sound. Correlation strongly depends on the choice of the parameter. This does not mean that in general correlation parameters are not suitable. For example, there is no doubt about the appropriateness of these parameters if comparing loudness model predictions with, e.g., magnitude estimation data. However, for the evaluation of loudness matching experiments the appropriate parameters must be found, and, if these parameters cannot be defined reliably, correlation-based parameters should be considered carefully. Therefore, it is recommended to focus on the reliable parameters ϵ and θ from the proposed set of indicators.

¹⁷ The mean is expected to lie between the lowest and the highest values of the 95% confidence interval with a probability of 95%.

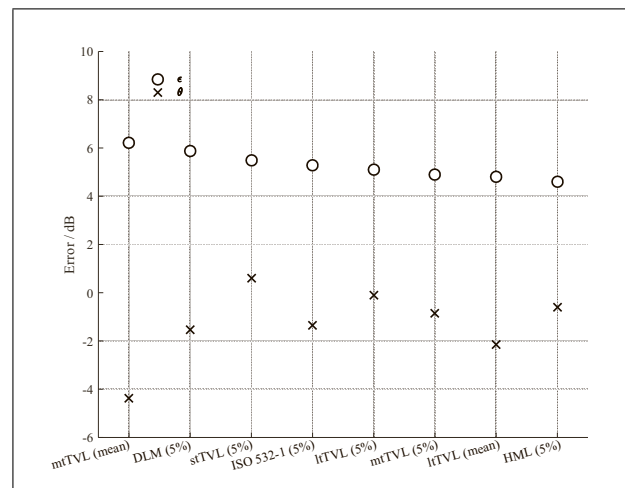


Figure 20. Root mean squared error ϵ , and the prediction bias (mean of prediction error) θ for Experiment 1.

Figure 20 shows the root mean squared error ϵ and the prediction bias θ for the different model predictions ordered from left to right starting with the highest value for ϵ . The hearing model loudness HML (5%) shows the lowest prediction error and a low prediction bias, followed by ltTVL (mean). The results from mtTVL (mean), DLM (5%) and stTVL (5%) show the highest prediction error. The results from ISO 532-1, ltTVL (5%) and mtTVL (5%) lie between.

3.5.2. Experiment 2: Replication of Experiment 1 using free-field equalized headphones

The open circles in Figure 21 show the mean level differences between test and reference sounds at the point of subjective equality and the corresponding 95% confidence intervals, based on averaged values for each participant and each test sound using three different starting values. The sounds are arranged in the same order as in Figure 19, only the sounds *impact wrench* and *screw wrench* were not used in the Experiment 2 (performed before the data of Experiment 1 were available). In general, the results of Experiment 2 differ only slightly from the results of Experiment 1 with respect to mean values (maximum difference 2.2 dB in the case of the sound *ratchet wheel*) and confidence intervals (maximum difference 1.5 dB in case of the sound *machine gun* and 1.4 dB in case of the sound *ratchet wheel*). Therefore, the general conclusions are the same as described above.

Intra-individual variability was considered by analyzing the three repetitions. The maximum difference observed between the results for three trials using the same sound with different starting levels was 14.5 dB (*machine gun*) for only one participant. This participant showed for other sounds a maximum difference of 1.5 dB. The lowest value observed for all participants and sounds was 0.5 dB. The mean value of this difference for all participants and all sounds was 4 dB. The highest value averaged over all participants was observed for the sounds *machine gun* (6.3 dB) and *ratchet wheel* (5.7 dB).

The root mean squared error ϵ , and the prediction bias θ for Experiment 2 are depicted in Figure 22. In general, ϵ is about 0.5 dB lower than in Experiment 1. HML (5%) again shows the lowest value, followed by ltTVL (mean). The sequence of order with respect to this quality measure is slightly changed compared to Experiment 1, but only for the algorithm showing larger prediction errors.

However, these quantities ϵ and θ are only quality indicators and not quality measures. Such a quality measure should be based on loudness, the parameter of interest. Measures like the root mean squared error ϵ_N , and the prediction bias (mean of prediction error) θ_N based on the loudness of the reference and test sounds at the point of subjective equality should be used instead, because these parameters consider the nonlinearity, the level-dependent loudness function.

Additionally, it should be taken into account that the true values of the experimental data are not known and can be only estimated. Therefore, a quality measure should consider the confidence intervals of the measured data. This can be achieved by the following procedure applied for each loudness model separately.

- For each sound of interest three new sounds are generated with a level according to
 - the lowest value of the confidence interval,
 - the mean value and
 - the highest value of the confidence interval at the point of equal loudness.
- The loudness of the adjusted sounds is predicted using the model of interest assuming that the loudness model performs a better prediction of loudness than simple level calculation. For example, a prediction error of, e. g., 6 dB in a loudness matching experiment using a sound at a moderate level where the loudness function is not as steep as at low or high levels should not be considered as a very large error, because it corresponds to a small or medium loudness difference. On the other hand, at low or high levels, a 6 dB level difference would lead to a larger loudness difference, and thus, should be considered as a larger error.
- Optionally, these calculated loudness values are normalized by the loudness of the corresponding reference sound in order to consider relative errors, as is performed in the following. The target line for optimal results is then a value of 1 (horizontal line as shown in Figure 23 for Experiments 1 and 2).

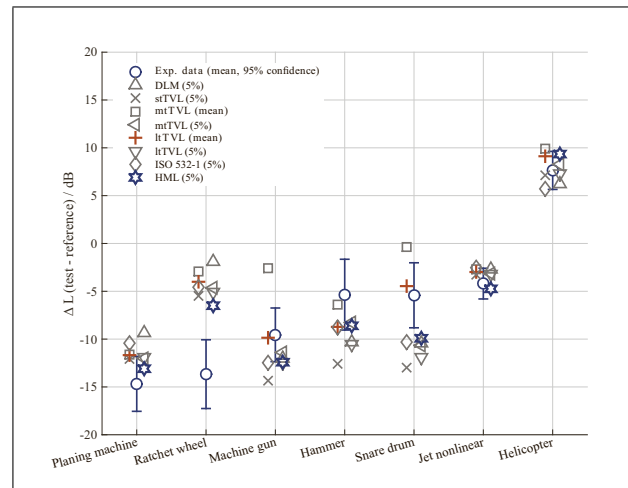


Figure 21. Results of loudness matching experiments using the indicated time-varying sounds and a stationary noise as reference (circles and bars: mean and 95% confidence intervals); the level differences at the point of subjective equality are given. Additionally, the predictions of various loudness models (based on N_5 - or mean values) are depicted.

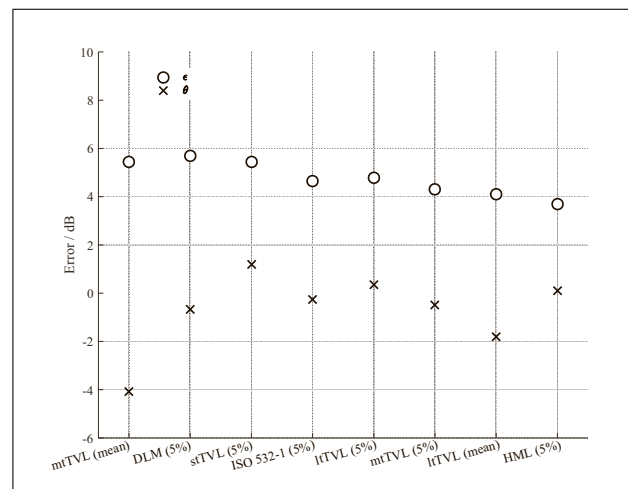


Figure 22. Root mean squared error ϵ , and the prediction bias (mean of prediction error) θ for Experiment 2.

- The root mean squared error σ of a distance measure between the predicted values and the confidence interval is taken as a quality measure. The distance measure is defined as
 - zero, if the prediction lies between the lowest and highest value of the 95% confidence interval, the desired case,
 - the difference between the lowest value of the 95% confidence interval and the prediction, if the prediction is smaller than lowest value of the 95% confidence interval,
 - the difference between the prediction and highest value of the 95% confidence interval, if the highest value of the 95% confidence interval is smaller than the prediction.

Table V. Quality indicators and measures of the different loudness models for Experiments 1 and 2.

		mtTVL (mean)	DLM (5%)	stTVL (5%)	ISO 532-1 (5%)	ltTVL (5%)	mtTVL (5%)	ltTVL (mean)	HML (5%)
Exp. 1	ϵ [dB]	6.2	5.9	5.5	5.3	5.1	4.9	4.8	4.6
	θ [dB]	-4.4	-1.5	0.6	-1.4	-0.1	-0.9	-2.2	-0.6
	ϵ_N [sone]	2.4	1.9	2.9	1.9	2.5	2.2	1.9	1.3
	θ_N [sone]	-1.7	-0.3	0.8	-0.3	0.3	-0.2	-0.8	-0.1
	σ [%]	12.2	9.3	10.4	8.3	8.1	6.8	6.7	3.8
Exp. 2	ϵ [dB]	5.4	5.7	5.4	4.7	4.8	4.3	4.1	3.7
	θ [dB]	-4.1	-0.7	1.2	-0.3	0.4	-0.5	-1.8	0.1
	ϵ_N [sone]	2.2	2.1	3.3	2.1	2.6	2.2	1.8	1.1
	θ_N [sone]	-1.7	0.1	1.1	0.0	0.5	0.0	-0.7	0.1
	σ [%]	11.0	11.1	14.1	9.1	8.9	6.3	5.0	2.9

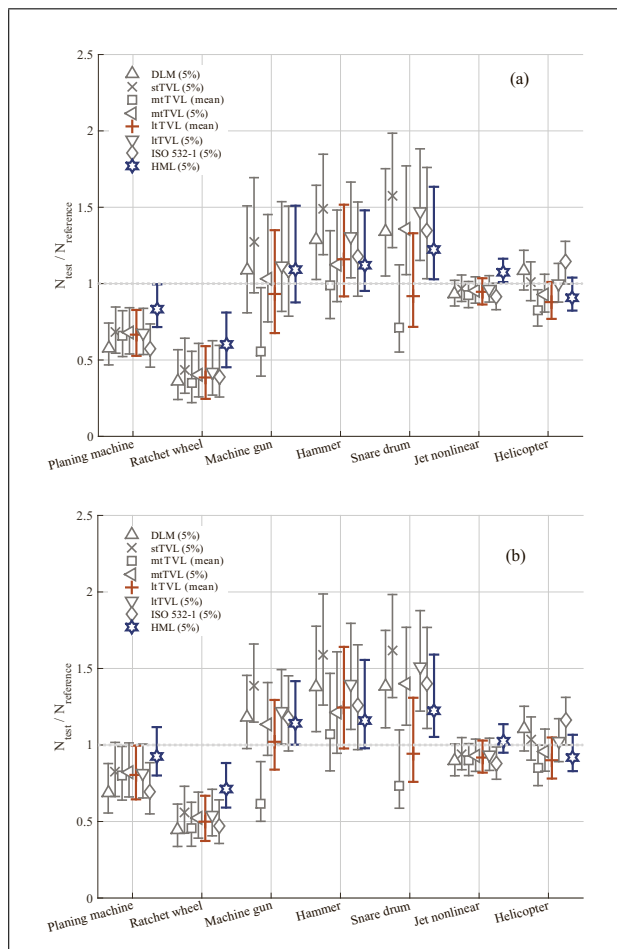


Figure 23. Loudness predictions based on 95% confidence intervals of adjusted levels for Experiment 1 (a) and 2 (b).

Figure 23 shows the results based on the loudness-based quality measures for both Experiments 1 and 2. In general HML (5%) shows smaller bars, caused by the stronger compression used. Figure 24 depicts the quality measure considering the confidence intervals of the measurements as described before. Table V lists all the numerical values of all quality indicators and measures.

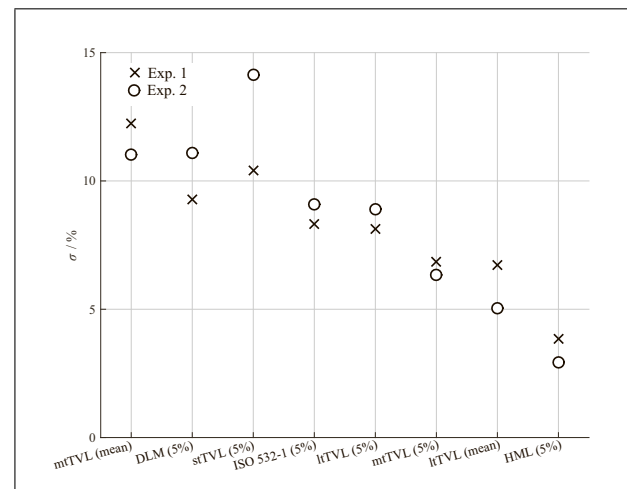


Figure 24. Prediction quality based on loudness deviation from 95% confidence intervals.

3.5.3. Experiment 3: Loudness of time-varying sounds with equal 3rd-octave spectra

Because the loudness predictions for some impulsive sounds do not match the ratings of normal-hearing listeners, the influence of specific signal properties of these sounds on the assessment of loudness was examined, starting with the sound *ratchet wheel* and other sounds provided by courtesy of Rennies, in order to investigate the reason why this sound was underestimated in loudness contrary to other impulsive sounds.

As described in sections 3.1 and 3.2 three different groups of three signals, e.g., *ratchet wheel* (original, reference sound), *white noise* (filtered, 3rd-octave spectrum of the sound *ratchet wheel*, test sound), *hammer* (filtered, 3rd-octave spectrum of the sound *ratchet wheel*, test sound) are used in this experiment. The corresponding loudness matching tests for this example are denoted by WN / L_T *Ratchet wheel* respectively $Hammer / L_T$ *Ratchet wheel* on the abscissa of Figure 25. The results of Experiment 3 (Figure 25) show that all models underestimate the loudness of the sounds *ratchet wheel* with modified 3rd-octave spectra (denoted by *Ratchet wheel* / L_T *WN* and *Ratchet wheel* / L_T *Hammer*), and some models also un-

Table VI. Quality measures of the different loudness models for Experiments 3 and 4.

		stTVL (5%)	mtTVL (mean)	ltTVL (5%)	ltTVL (mean)	mtTVL (5%)	DLM (5%)	ISO 532-1 (5%)	HML (5%)
Exp. 3	ϵ_N [sone]	2.3	1.8	1.8	1.7	1.6	1.6	1.3	0.6
	θ_N [sone]	0.5	0.3	0.4	0.3	0.4	0.4	0.4	0.3
	σ [%]	5.0	8.6	2.5	4.9	3.1	7.4	3.1	0.0
Exp. 4	ϵ_N [sone]	3.4	3.0	3.0	2.9	2.8	2.7	2.5	1.4
	θ_N [sone]	-0.2	-2.1	-0.8	-1.4	-1.3	-1.1	-1.2	-0.5
	σ [%]	22.5	23.9	20.7	22.8	19.1	27.2	22.5	10.2

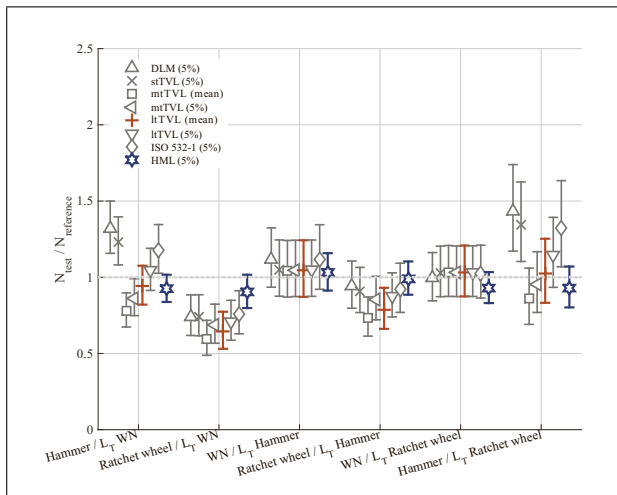


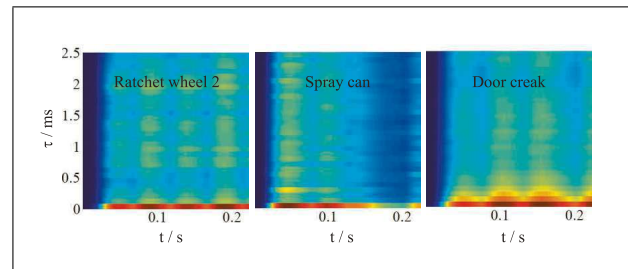
Figure 25. Loudness predictions based on 95% confidence intervals of adjusted levels for Experiment 3.

derestimate the original sound *ratchet wheel* as reference stimulus (denoted by *Hammer / L_T Ratchet wheel*). The hearing model of Sottek predicts the results of this experiment within the 95% confidence intervals of the adjusted levels. The quality measures of the different loudness models for this experiment are listed in Table VI, sorted with respect to the root mean squared error ϵ_N .

3.5.4. Experiment 4: Loudness of new time-varying sounds

Figure 12 illustrates the fact that the summed ACF of the sound *ratchet wheel* has different features than the sound *hammer*. It was verified that the decay of the ACF is much steeper for the sound *ratchet wheel* than for all other sounds used in Experiment 1. After observing this special feature of the sound *ratchet wheel*, new test sounds were searched for with similar properties. As examples, the sound *ratchet wheel 2* (another cut of the recording as indicated in Table III), the sound *spray can* (rattling sound), and the sound *door creak* were analyzed.

On the one hand, Figure 26 shows that the results for the sounds *ratchet wheel 2* and *spray can* have one key aspect in common with the result of the first sound *ratchet wheel*: 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 sound *door creak* is very similar to

Figure 26. Sum of the autocorrelation functions of all critical bands as a function of time t and lag τ for the sounds *ratchet wheel 2*, *spray can* (rattling sound), and *door creak*.

that of the sound *hammer*. These sounds are less impulsive and have stronger low frequency contents.

It is hypothesized that the loudness of the sounds *ratchet wheel 2* and *spray can* are underestimated by the existing loudness models as in the case of the sound *ratchet wheel*. This should not be the case for the sound *door creak*. Figure 27 supports this assumption.

Figure 27 also shows results achieved by the hearing model loudness approach but without any consideration of the ACF vs. lag functions. The results of this experiment are predicted in general better by the hearing model compared to the other models. The quality measures of the different loudness models for this experiment are listed in Table VI, sorted with respect to root mean squared error ϵ_N .

4. Discussion

The first two experiments were considered in order to compare results of loudness matching experiments performed in two different studies with different software and hardware, but with the same methodology. In general, the results were very similar, as described above. The method to evaluate the prediction quality was different between the two studies; whereas Rennie *et al.* [3] are concentrating more on correlation values than on root mean squared errors of level differences at equal loudness, the focus of the present study lies on prediction measures based on loudness, the parameter of interest. The main conclusion of Rennie *et al.* that ltTVL (mean) performs the best among the considered models, is based mainly on correlation (R^2 and ρ_S). Looking at the root mean squared error ϵ the difference between DIN 45631/A1 and ltTVL is, for the con-

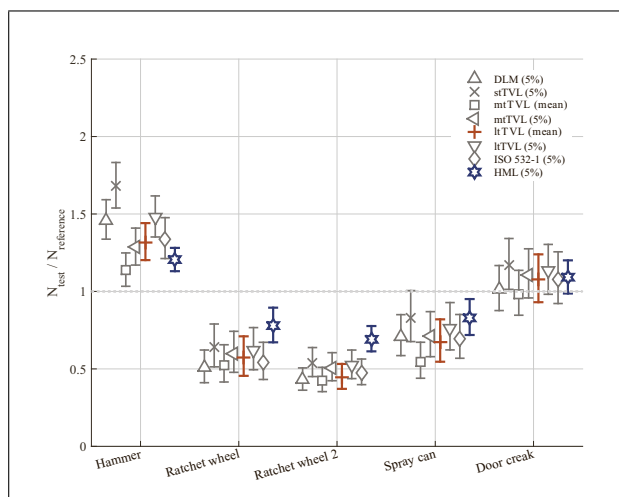


Figure 27. Loudness predictions based on 95% confidence intervals of adjusted levels for Experiment 4.

sidered Experiment 1, still in favor of ltTVL (by 0.7 dB)¹⁸, but for the other experiments from Rennie *et al.* [3] (not covered in this paper) it is in favor of DIN 45631/A1 (by 0.4 dB and 0.8 dB). The hearing model approach to time-varying loudness HML (5%) performed (even for non-recommended correlation parameters) the best, followed by ltTVL (mean) and mtTVL (5%). DLM (5%), stTVL (5%) and mtTVL (mean) were less accurate.

Experiment 3 covered sounds with equal 3rd-octave spectra but different time structures of reference and test signals where stationary loudness models would show almost no loudness difference for the unmodified sounds. The root mean squared error with respect to loudness ϵ_N lies between 0.6 sone for HML (5%) and 2.3 sone for stTVL (5%) without considering the confidence intervals of the measurements. It should be considered that the different models are using different nonlinearities and, thus, provide a different loudness scaling. Therefore, relative errors may be better comparable: the introduced measure σ shows the best value for HML (5%), followed by ltTVL (5%), mtTVL (5%) and ISO 532-1 (5%); mtTVL (mean), DLM (5%), stTVL (5%) and ltTVL (mean) were less accurate. In general, the prediction quality was only slightly better than for Experiments 1 and 2 where spectro-temporal characteristics were varied. Largest deviations were found for the sound *hammer* with the same 3rd-octave spectra as the sound *ratchet wheel* adjusted in loudness to the sound *ratchet wheel*, but small deviations for the filtered WN adjusted in loudness to the sound *ratchet wheel*. This indicates that for the sound *ratchet wheel* the loudness depends strongly on both spectral and temporal characteristics, as also concluded by Rennie *et al.* [3].

Experiment 4 was similar to Experiments 1 and 2 but with new time-varying signals. As described above, the motivation for this experiment was to find sounds with

similar characteristics as the sound *ratchet wheel* in order to investigate if other features of the ACF for the hearing model loudness of Sottek could be used to improve loudness calculation. The first results are very promising for proceeding further in this direction. The root mean squared error with respect to loudness ϵ_N lies between 1.4 sone for HML (5%) (without any consideration of the ACF vs. lag functions) and 3.4 sone for stTVL (5%). ISO 532-1 (5%) shows, as for Experiment 3, the second lowest deviation of 2.5 sone. The introduced measure σ shows values between 10.2% for HML (5%) and 27.2% for DLM (5%). The results for mtTVL (5%) show with 19.1% the second lowest value.

In this study three different versions of the TVL model were considered using different temporal integration: the stTVL method with only one temporal integration stage and two other versions using a second temporal integration stage with different release time constants: 2000 ms (ltTVL) and 200 ms (mtTVL). It could be shown that a change of the release time constant by one order of magnitude yields almost the same results for Experiment 1 to 3 and even better results for Experiment 4 where sounds with special temporal features were evaluated, provided that ltTVL (mean) and mtTVL (5%) are used to derive overall loudness. The very long release time constant of the ltTVL method seems to be essential for good loudness prediction if averaging is applied, yielding always better results than the mtTVL (mean) method. But if percentile loudness values are needed, e.g., to be comparable to other models based on *short-term loudness*, the mtTVL method should be used. The stTVL method using the shortest time constants showed the poorest results among the three TVL versions.

5. Conclusion

The existing loudness models cannot always reliably predict the perceived loudness of sounds, such as in the case of impulsive sounds. For some of the investigated impulsive sounds, loudness is strongly underestimated (e. g., the sound *ratchet wheel*) while for other sounds, loudness is overestimated (e. g., the sound *hammer*). The presented hearing model approach to time-varying loudness HML (5%) performs better than other existing loudness models, followed by ltTVL (mean), the introduced mtTVL (5%) as a modified version of the ltTVL model (with a release time constant of 200 ms for the second cascade of temporal integration instead of 2000 ms) and the new loudness standard ISO 532-1 (5%) (based on DIN 45631/A1).

It could be found that sounds with adapted 3rd-octave spectra but different time structures may have significantly different perceived loudness, although many existing models (with the exception of the HML) do not predict these differences.

An overview of a theoretical system model of the monaural signal processing in the human auditory system, introduced by Sottek from 1989 to 1993, was given which explains and describes not only special aspects of

¹⁸ In this study, using the TVL software from <http://hearing.psychol.cam.ac.uk/Demos/demos.html>, only a difference of 0.5 dB could be found.

sound perception, but also many psychoacoustic phenomena. The focus was placed on a functional scheme to calculate the excitation pattern on the basilar membrane and to model the nonlinear behavior with respect to just-noticeable amplitude differences and variations.

The described hearing model, characterized especially by the application of the nonlinearity and the steeper curve progression at higher levels, leads to a significant improvement of the prediction quality for several loudness matching experiments using synthetic and technical sounds.

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 a further improved loudness prediction of impulsive sounds.

Acknowledgement

The author wishes to thank Farzaneh Parvizian, Tobias Moll and Valeriya Tarasova for their support during the experiments and two anonymous reviewers and the editor for valuable comments on earlier versions of the manuscript.

References

- [1] H. Fastl, E. Zwicker: *Psychoacoustics. Facts and models.* Springer Verlag, Berlin, Heidelberg, New York, 2006.
- [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* **19** (2013) 050189.
- [3] J. Rennies, M. Wächtler, J. Hots, J. L. Verhey: Spectro-temporal characteristics affecting the loudness of technical sounds: data and model predictions. *Acta Acustica united with Acustica* **101** (2015) 1145–1156.
- [4] R. Sottek: Loudness models applied to technical sounds. *Noise-Con 2010*, Baltimore, 2010.
- [5] 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** (2013) EL334–EL339.
- [6] 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** (2014) 1313–1320.
- [7] J. Chalupper, H. Fastl: Dynamic loudness model (DLM) for normal and hearing-impaired listeners. *Acta Acustica united with Acustica* **88** (2002) 378–386.
- [8] J. Rennies, J. L. Verhey, J. L. Chalupper, H. Fastl: Modeling temporal effects of spectral loudness summation. *Acta Acustica united with Acustica* **95** (2009) 1112–1122.
- [9] 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.
- [10] ISO 532-1: Acoustics - Methods for calculating loudness. Part 1: Zwicker method. to be published, 2016.
- [11] ISO 532:1975: Acoustics - Methods for calculating loudness level. 1975.
- [12] E. Zwicker: Procedure for calculating loudness of temporally variable sounds. *J. Acoust. Soc. Am.* **62** (1977) 675–682.
- [13] 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.
- [14] R. Sottek: Progresses in standardizing loudness of time-variant sounds. *Proc. AIA-DAGA 2013*, Merano, 2013, 1125–1128.
- [15] R. Sottek: *Modelle zur Signalverarbeitung im menschlichen Gehör.* Dissertation, RWTH Aachen, 1993.
- [16] R. Sottek, F. Kamp, A. Fiebig: A new hearing model approach to tonality. *Proc. Internoise*, Innsbruck, 2013.
- [17] R. Sottek: Progress in calculating tonality of technical sounds. *Proc. Internoise*, Melbourne, 2014.
- [18] E. Zwicker, R. Feldtkeller: *Das Ohr als Nachrichtenempfänger.* 2nd edition. S. Hirzel Verlag, Stuttgart, 1967.
- [19] B. C. J. Moore, B. R. Glasberg: Suggested formulae for calculating auditory filter bandwidths and excitation patterns. *J. Acoust. Soc. Am.* **74** (1983) 750–753.
- [20] B. C. J. Moore, B. R. Glasberg: The role of frequency selectivity in the perception of loudness, pitch and time. – In: *Frequency Selectivity in Hearing.* B. J. C. Moore (ed.). Academic Press, London, 1986, chapter 5, 251–308.
- [21] E. Zwicker, G. Flottorp, S. S. Stevens: Critical band width in loudness summation. *J. Acoust. Soc. Am.* **29** (1957) 548–557.
- [22] E. Zwicker: Subdivision of the audible frequency range into critical bands (Frequenzgruppen). *J. Acoust. Soc. Am.* **33** (1961) 248–249.
- [23] E. Zwicker: Über psychologische und methodische Grundlagen der Lautheit. *Acustica* **8** (1958) suppl. 1, 237–258.
- [24] S. S. Stevens: The measurement of loudness. *J. Acoust. Soc. Am.* **27** (1955) 815–829.
- [25] J. Hellbrück: Loudness scaling: Ratio scales, rating scales, intersubject variability and some other problems. – In: *Contributions to Psychological Acoustics. Results of the Fifth Oldenburg Symposium on Psychological Acoustics.* A. Schick, J. Hellbrück, R. Weber (eds.). BIS University of Oldenburg, Oldenburg, 1991, 110–133.
- [26] O. Heller: Oriented category scaling of loudness and speech audiometric validation. – In: *Contributions to Psychological Acoustics. Results of the Fifth Oldenburg Symposium on Psychological Acoustics.* A. Schick, J. Hellbrück, R. Weber (eds.). BIS University of Oldenburg, Oldenburg, 1991, 135–159.
- [27] W. R. Garner: A technique and a scale for loudness measurement. *J. Acoust. Soc. Am.* **26** (1954) 73–88.
- [28] P. M. Sellick, R. Patuzzi, B. M. Johnston: Measurement of the basilar motion in the guinea pig using the Mössbauer technique. *J. Acoust. Soc. Am.* **72** (1982) 131–141.
- [29] E. Schorer: Vergleich eben erkennbarer Unterschiede und Variationen der Frequenz und Amplitude von Schallen. *Acustica* **68** (1989) 183–199.
- [30] T. Bierbaums, R. Sottek: Modellierung der zeitvarianten Lautheit mit einem Gehörmodell. *Proc. DAGA 2012*, Darmstadt, 2012, 591–592.
- [31] R. Sottek: Improvements in calculating the loudness of time-varying sounds. *Proc. Internoise*, Melbourne, 2014.
- [32] S. Buus, M. Florentine: Modifications to the power function for loudness. – In: *Fechner Day 2001. Proceedings of the 17th Annual Meeting of the International Society for*

- Psychophysics. E. Summerfield, R. Kompass, T. Lachmann (eds.). Pabst, Berlin, 2001, 236–241.
- [33] M. Epstein, M. Florentine: A test of the equal-loudness-ratio hypothesis using cross-modality matching functions. *J. Acoust. Soc. Am.* **118** (2005) 907–913.
- [34] R. Sottek: Gehörgerechte Rauigkeitsberechnung. *Proc. DAGA 1994, Dresden, 1994*, 1197–1200.
- [35] R. Sottek, P. Vranken, H.-J. Kaiser: Anwendung der gehörgerechten Rauigkeitsberechnung. *Proc. DAGA 1994, Dresden, 1994*, 1201–1204.
- [36] B. C. J. Moore, B. R. Glasberg, C. J. Plack, A. K. Biswas: The shape of the ear's temporal window. *J. Acoust. Soc. Am.* **83** (1988) 1102–1116.
- [37] A. Vogel: Ein gemeinsames Funktionsschema zur Beschreibung der Lautheit und der Rauigkeit. *Biological Cybernetics* **18** (1975) 31–40.
- [38] D. Maiwald: Die Berechnung von Modulationsschwellen mit Hilfe eines Funktionsschemas. *Acustica* **18** (1967) 193–207.
- [39] J. C. R. Licklider: A duplex theory of pitch perception. *Cellular and Molecular Life Sciences* **7** (1951) 128–134.
- [40] M. Florentine, M. Epstein: To honor Stevens and repeal his law (for the auditory system). – In: Fechner Day 2006. Proceedings of the 22nd Annual Meeting of the International Society for Psychophysics. D. E. Kornbrot, R. M. Msetfi, A. W. MacRae (eds.). St. Albans, UK, 2006, 37–42.
- [41] J. Houts, J. Rennie, J. L. Verhey: Modeling temporal integration of loudness. *Acta Acustica united with Acustica* **100** (2014) 184–187.
- [42] T. Poulsen: Loudness of tone pulses in a free field. *J. Acoust. Soc. Am.* **69** (1981) 1786–1790.
- [43] S. D. Ewert: AFC - a modular framework for running psychoacoustic experiments and computational perception models. *Proc. AIA-DAGA 2013, Merano, 2013*, 1326–1329.
- [44] J. L. Verhey: Psychoacoustics of spectro-temporal effects in masking and loudness perception. Phd thesis, University of Oldenburg, Germany, 1999.