Considerations in modeling the nonlinearity of human hearing for loudness perception

Julian Becker[1], Roland Sottek[1], Tao-Li Liu[1], Moritz Bender[1],

[1] *HEAD acoustics GmbH, 52134 Herzogenrath, Germany*

Introduction

The relationship between human loudness perception and sound pressure level is nonlinear. In the past, several experiments have been conducted to investigate this behavior. A simplified model for this relationship is the assumption of a factor of two in loudness perception for a level difference of 10 dB for pure tones with a loudness level above 40 phon. This assumption is used, for example, in the Zwicker Loudness Model, which is standardized in ISO 532-1 [1], and in the Moore-Glasberg-Schlittenlacher Loudness Model, standardized in ISO 532-3 [2]. However, the Sottek Hearing Model Loudness, standardized in ECMA-418-2 [3], uses a more elaborate approximation of the nonlinear perception of humans, which results in a more accurate estimate of loudness compared to the results of listening tests. Consequently, the estimated loudness from this method must differ from the other models that use simplified approximations.

In this paper, we review the existing data on the nonlinearity of human hearing in terms of loudness perception and analyze the effects of different approximations in the various loudness standards ISO 532-1, ISO 532-3, and ECMA-418-2. The results show that the consequence of a more precise approach to human nonlinear perception leads to significant differences in the estimated loudness in sone.

Review of history

The unit sone for a linear scale of human loudness perception was introduced by Stevens [4] almost 90 years ago. Stevens suggested the unit of a loudness scale should be the loudness of a pure tone with a frequency of 1000 Hz and a sound pressure level (SPL) of 40 dB. Stevens also suggested that the relationship between loudness in sone and sound pressure level can be described by a power law. From this power law, it was concluded that for every increase of 10 phon the perceived loudness in sone doubles.

The approximative correctness of the power law was validated in several experiments, mainly using magnitude estimation or methods of constant stimuli (e.g. “half as loud”, “double as loud” using either pure tones or noise signals. An overview of several experiments of those early days is given by Stevens in [5]. From these early experiments, it was already indicated that the simple assumption of a power law might not be sufficient to explain the relationship between loudness level and loudness. Robinson concluded from an experiment using the method of constant stimuli with doubling and halving loudness [6]: “There is a significant difference between the scales relating to increase and decrease of sensation respectively, especially at low levels of loudness. The interval on the phon scale for a given loudness ratio is approximately constant for *decreasing* sensation, but for loudness *magnification* it decreases as the absolute loudness is raised.” Even though this indicates that a simple power law might not be sufficient, Robinson also concluded that his results were in fair agreement with previous work and could also be described by a power law.

Zwicker [7] also studied the nonlinearity of human hearing by performing listening tests of double and half loudness perception. Zwicker’s results for 1000 Hz pure tones were in agreement with the studies of Stevens and Robinson. Zwicker also studied noise sounds and concluded that the loudness of 1000 Hz tones is not the basic element for the loudness of more complex sounds.

Hellman and Zwislocki [8] indicated that there was a need to modify the simple power law for low levels close to the hearing threshold where the function is much steeper than described by the power law. This finally led to the modification that the power law is only used for loudness levels above 40 phon and modified for lower loudness levels. This approximation is still widely used today.

Zwicker also developed a similar function for describing the nonlinear behavior of human hearing for his loudness model, which was standardized in the German standard DIN 45631 [9]. This method was also used as the basis for the ISO 532-1 standard [1].

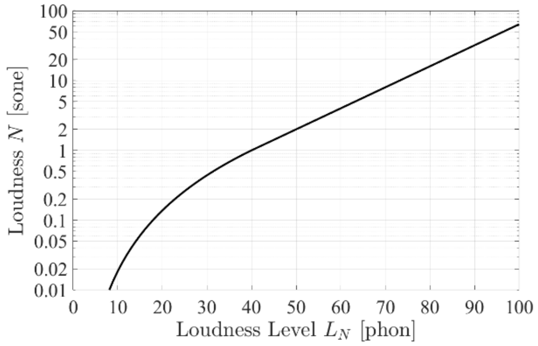

Figure 1: Transformation of loudness level in phons to loudness in sones according to the Zwicker model.

Figure 1 shows the relationship between loudness level in phons and loudness in sones according to the Zwicker model. This is one of the most widely used approximations of the nonlinearity of human hearing until today.

Moore and Glasberg [10] formulated a revision of Zwicker’s loudness model. While this revision suggests a different approach in calculating loudness than Zwicker’s model, it does not propose any significant change to the modeling of the nonlinearity of human hearing. The Moore and Glasberg loudness model was used as the basis of the ISO 532-3 [2] standard.

The relationship of loudness and loudness level is a topic of ongoing research. This research often shows that the simplification of using a power law for all loudness levels above 40 phon is not sufficient and therefore a more detailed approximation is desired. Florentine and Epstein [11] stated: “At low levels, loudness grows more rapidly. At moderate levels, loudness shows compressive growth. At high levels, the rate of loudness growth again increases." To account for these variations, they suggested to use an "inflected exponential function" (INEX) that replaces the constant slope of the simple power law with a continuous polynomial, allowing the slope to vary slowly with level. A particular implementation of such an INEX loudness function is given in [12].

A similar approach was taken by Sottek [13] who suggested a power law with piecewise-defined exponents. This function does consider the steeper slope at higher levels as described by Florentine and Epstein, but also takes into account newer studies, also considering other types of signals (e.g. narrow-band noise) [14] and other procedures such as consideration of just noticeable differences [15]. This approach is also used in the ECMA-418-2 standard to calculate loudness based on the Sottek Hearing Model, but also as the basis for other psychoacoustic measures such as tonality and roughness.

While this brief review of the history of research on the nonlinearity of human hearing does not claim to be complete, it shows that the topic is rather complex and the understanding of loudness perception has evolved over time. It also shows that the simple power law which is still widely used today might be too simplified and that more accurate models have been developed.

Standardized Loudness Methods

Introduction

Several algorithms have been developed to model human perception of loudness. In this paper, the algorithms standardized in ISO 532-1 (Zwicker method), ISO 532-3 (Moore-Glasberg-Schlittenlacher method) and ECMA-418-2 (Sottek Hearing Model) are compared.

ISO 532-1

ISO 532-1 is an international standard for measuring perceived loudness using the Zwicker method. It provides a computational model to estimate how humans perceive the loudness of stationary and time-varying sounds. The method incorporates critical band analysis, which divides the frequency spectrum into bands that approximate the human ear's resolution, and accounts for temporal masking effects.

Regarding the nonlinearity of human hearing, the method uses the simplified power law model by transforming phons to sones according to the equation

$$N = 2^{\frac{\left(\frac{L_N}{\text{phon}}-40\right)}{10}} \qquad \text{[sone]} \qquad (1)$$

for all loudness levels $L_N > 40$ phon. For lower loudness levels, the relationship between phons and sones is modeled as

$$L_N = 40\left(\frac{N}{\text{sone}} + 0.0005\right)^{0.35}, \qquad \text{[phon]} \qquad (2)$$

considering the steeper function for lower levels.

ISO 532-3

ISO 532-3 is an international standard for measuring perceived loudness using the Moore-Glasberg-Schlittenlacher method, a computational model based on the human auditory system. This method provides an approach to loudness measurement, particularly for complex and non-stationary sounds.

A similar method for stationary sounds is described in ISO 532-2. ISO 532-3 accounts for factors like frequency resolution, masking effects, and temporal integration.

The relationship between phons and sones is not directly expressed in form of an equation in the standard, but a table is given which specifies this relationship.

Table 1: Relationship between loudness level in phons and loudness in sones in ISO 532-3

Loudness level [phon]	Calculated loudness [sone]	Loudness level [phon]	Calculated loudness [sone]
0	0.001	60	4.11
10	0.025	70	7.88
20	0.137	80	15.2
30	0.42	90	30.9
40	1.00	100	64.6
50	2.09		

While this relationship does not exactly follow the power law as used in ISO 532-3, it can be observed that the relationship is very similar. Where the power law used in ISO 532-1 results in a loudness of 64 sone for 100 phon, ISO 532-3 results in a loudness of 64.6 sone. For levels below 40 phon ISO 532-3 also shows a more compressive behavior.

ECMA-418-2

ECMA-418-2 is a standard developed by ECMA International that describes the Sottek Hearing Model and psychoacoustic metrics dependent on the hearing model. The hearing model, which is used as the basis for several psychoacoustic metrics, expresses specific loudness for stationary and time-varying sounds. While the ISO methods do not differentiate between the calculation of the loudness of noise and tones that fall into the same critical band, the ECMA-418-2 method separates noise and tonal loudness, enabling the algorithm to treat these different signals individually, as they are also perceived differently by human listeners [16]. The estimated perceived loudness is given in sone$_{\text{HMS}}$. The subscript HMS (Hearing Model of Sottek) indicates that this sone-unit differs from the classical definition. The main reason for this is the different relationship between phons and sones$_{\text{HMS}}$. In the standard, the transformation to a first estimate of the specific loudness from sound pressure values in calculated as

$$\tilde{N}' = c_N \cdot \frac{\tilde{p}}{\tilde{p}_0} \cdot \prod_{i=1}^{8}\left(1+\left(\frac{\tilde{p}}{\tilde{p}_{t_i}}\right)^{1.5}\right)^{\frac{v_i - v_{i-1}}{1.5}}, \qquad (3)$$

where $\tilde{N}'$ is the specific loudness without consideration of the threshold in quiet. The variable $\tilde{p}$ is the root-mean-square of preprocessed (e.g. bandpass-filtered and filtered with outer- and middle/inner ear filters) sound pressure $\tilde{p}_{t_i}$ are thresholds in Pa, $\tilde{p}_0 = 20$ µPa is the reference sound pressure and c_N is a calibration factor. The exponents v_i are defined depending on the thresholds as shown in Table 2.

While the effect of this nonlinearity might not be intuitively clear, it is obvious that the partial adaptation of the exponents in eight segments allows a much closer approximation of the nonlinearity of human hearing than a simple power law with one fixed exponent.

Table 2: Thresholds and exponents for the ECMA-418-2 nonlinearity function (Eq. (3))

i	1	2	3	4
$20\log_{10}\left(\frac{\tilde{p}_{t_i}}{\tilde{p}_0}\right)$ [dB]	15	25	35	45
v_i	0.6602	0.0864	0.6384	0.0328
i	5	6	7	8
$20\log_{10}\left(\frac{\tilde{p}_{t_i}}{\tilde{p}_0}\right)$ [dB]	55	65	75	85
v_i	0.4068	0.2082	0.3994	0.6434

Comparing the Output with Experimental Data

Used data

In this chapter, the results of the three standards ISO 532-1, ISO 532-3 and ECMA 418-2 are compared with the results of experimental data. The experimental data are taken from Zwicker's experiments on loudness perception [7], which were the basis for the development of the Zwicker loudness model. One of the main contributions regarding the nonlinearity of human hearing in [7] are Zwicker's experiments on the perception of double and half loudness as shown in Figure 2. The figure shows the mean values and the lower and upper quartiles of the results of Zwicker's experiments. The theoretical values for double and half loudness perception according to the Zwicker model as described in Equation (1) and (2) are additionally shown in Figure 2.

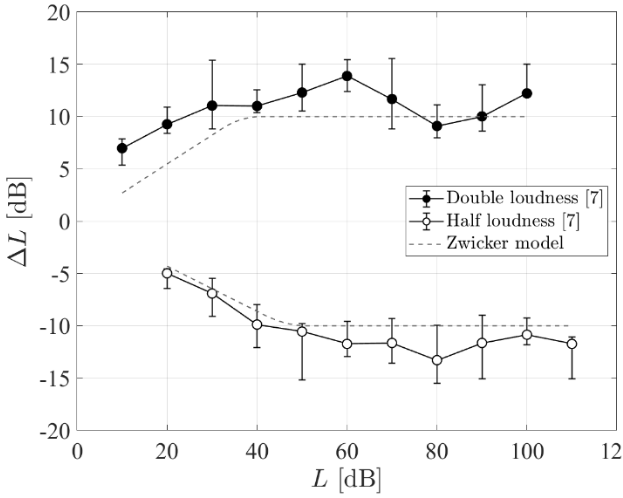

Figure 2: Matching of double/half loudness experiment of Zwicker [7] and modeling using the Zwicker method (used in ISO 532-1).

Interestingly, the Zwicker model is systematically below the experimental data. This indicates an overestimation of loudness for this model in comparison to the experimental results for loudness levels above 40 phon and an underestimation below 40 phon. Through cumulative effects this difference should increase for loudness levels further from 40 phon. The representation in Figure 2 does not directly show how large these deviations are from the expected loudness of the experimental data to the one of the model in sone. To visualize these effects, we transformed the results of the experimental data in Figure 2 into a direct conversion from phons to sones, as displayed in Figure 3. The black markers correspond to the exact mean values of the double loudness perception in Figure 2. For example, in Figure 2, a doubling of loudness perception is shown at a mean ΔL of 11.01 dB. Accordingly, in Figure 3, the corresponding marker is put at a loudness level of 51.01 phon and a corresponding loudness of 2 sone. Between the black markers, a piecewise exponential function was used to interpolate the continuous values.

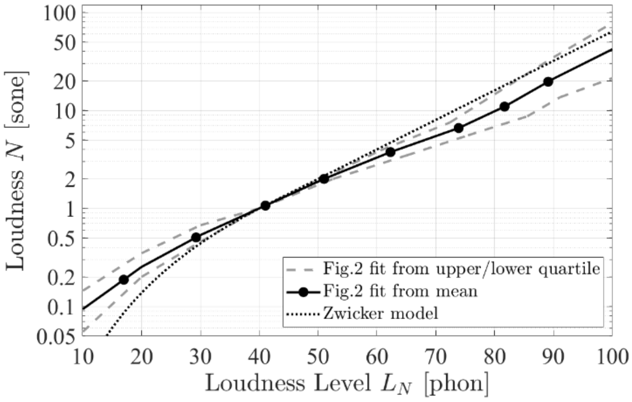

Figure 3: Fitted curves from data in Figure 2 (double loudness).

The same procedure was performed for the upper and lower quartiles of the double loudness experiment in Figure 2. The resulting continuous results are shown as dashed lines in Figure 3. This can be interpreted as a corridor within which the assumption of a linear perceptual scale is correct according to Zwicker's results of double loudness perception. Additionally, Zwicker's loudness model as described by Equations (1) and (2) is shown as dotted line in Figure 3. The expected underestimation for lower loudness levels and overestimation for higher loudness levels can be clearly observed. For example, for a loudness level of 80 phon, the expected loudness in sone from the fit to the mean value of the experimental data would be 9.8 sone, whereas the Zwicker model results in a value of 16 sone.

To compare the results of the loudness standards ISO 532-1, ISO 532-3 and ECMA 418-2 with these experimental results, the loudness in sone was calculated with each of these standard for 1000 Hz sinusoids with levels ranging from 10 dB to 80 dB with a step size of 1 dB.

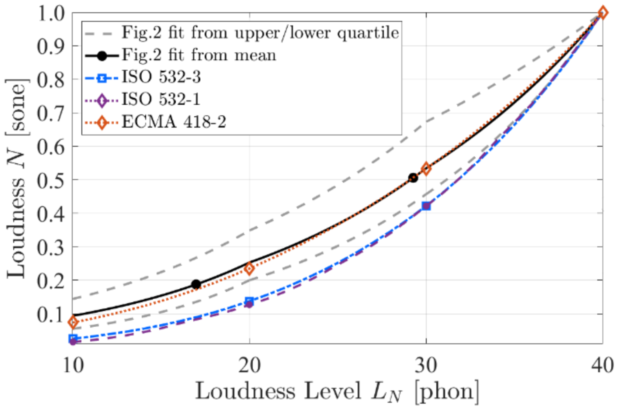

Figure 4: Fitted curves with results for ISO 532-1, ISO 532-3 and ECMA 418-2 (10-40 phon).

For ease of interpretation, the results are shown in two different figures: Results from 10 to 40 phon are shown in Figure 4, results for loudness levels above 40 phon are shown in Figure 5. In both figures, the loudness is shown on a linear scale for a more intuitive interpretation of the deviations.

In Figure 4, it can be observed that the results of the ISO 532-1 are consistently below the lower dashed line corresponding to the quartile range of the experimental data. This is due to the underestimation of loudness of the Zwicker model in this range as mentioned above. The results of the ISO 532-3 standard are in a very similar range. The results of the ECMA 418-2 are very close to the fit to the mean results of the experimental data in this loudness level range.

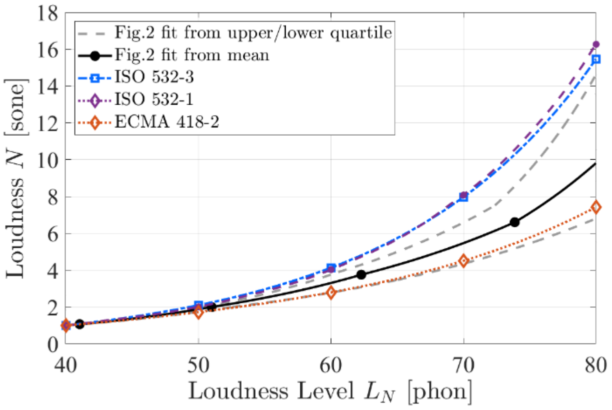

Figure 5: Fitted curves with results for ISO 532-1, ISO 532-3 and ECMA 418-2 (40-80 phon).

In Figure 5, the result for loudness levels from 40 to 80 phon are displayed. Again, ISO 532-1 and 532-3 lead to very similar results. Those results are higher than to be expected from Zwicker's experimental data. The ECMA-418-2 standard leads to significantly lower results. However, these lower estimates of loudness still fall into the corridor of the fitted quartiles of the experimental data.

In conclusion, the results show that the ISO 532-1 and the ISO 532-3 standards use a very similar model of the nonlinearity of human loudness perception. The ECMA 418-2 standard uses a very different approach which results in a very different estimation of loudness. This difference is also reflected in the usage of a different unit: sone$_{HMS}$ instead of sone. The comparison with Zwicker's experimental data shows that the different approach of the ECMA 418-2 standard fits well to this established data set about the nonlinearity of human loudness perception.

Conclusion

In this paper, we have compared the modeling of the nonlinearity of human loudness perception of the three loudness standards ISO 532-1, ISO 532-3 and ECMA 418-2. We compared the results of the mapping from loudness level in phons to loudness in sones of these three standards and compared them with historical experimental data collected by Zwicker.

The results show that the two ISO standards lead to rather similar results, while the estimated loudness of the ECMA 418-2 standard differs significantly. The comparison with Zwicker's experimental data shows that the different approach used in the Sottek Hearing Model, which is the basis of the ECMA 418-2 standard fits well to the data. The results show that it is important to understand that loudness calculated with different methods is not comparable and results obtained from different standards cannot be directly compared.

The results presented in this paper do not provide a complete picture of the effects of the modeling of the nonlinearity of human loudness perception, but add some more insight to other publications, such as [16], where the performance of loudness standards was studied for narrowband noise. The results of the different loudness standards in relation to the equal loudness contours were studied in [17].

Literature

[1] ISO 532-1: Acoustics – Methods for calculating loudness, Part 1: Zwicker method. (2017)

[2] ISO 532-3, Acoustics – Methods for calculating loudness, Part 3: Moore-Glasberg-Schlittenlacher method. (2023)

[3] ECMA-418-2: Psychoacoustic metrics for ITT equipment - Part 2: Methods for describing human perception based on the Sottek Hearing Model, third edition. (2024)

[4] Stevens, S. S.: A scale for the measurement of a psychological magnitude: loudness. Psychological Review, 43(5), 405–416. (1936)

[5] Stevens, S. S.: The measurement of loudness. The Journal of the Acoustical Society of America 27.5 815-829. (1955)

[6] Robinson, D. W.: "The relation between the sone and phon scales of loudness. Acta Acustica united with Acustica 3.5 344-358. (1953)

[7] Zwicker, E.: Über psychologische und methodische Grundlagen der Lautheit. Acta Acustica united with Acustica 8.4 237-258. (1958)

[8] Hellman, R.P. and Zwislocki, J.: Some factors affecting the estimation of loudness. The Journal of the Acoustical Society of America, 33(5), 687-694. (1961)

[9] DIN 45631: Berechnung des Lautstärkepegels aus dem Geräuschspektrum, Verfahren nach E. Zwicker. (1967)

[10] Moore, B. C. J. and Glasberg, B. R.: A revision of Zwicker's loudness model. Acta Acustica united with Acustica 82.2 335-345. (1996)

[11] Florentine, M. and Epstein, M.: To honor Stevens and repeal his law (for the auditory system). Fechner Day, 37-42. (2006)

[12] Epstein, M. and Marozeau, J.: Loudness and intensity coding. The Oxford Handbook of Auditory Science–Hearing, 3, 45-69. (2010)

[13] Sottek, R.: A hearing model approach to time-varying loudness. Acta Acustica united with Acustica, 102(4), 725-744. (2016)

[14] Bierbaums, T. and Sottek, R.: Modellierung der zeitvarianten Lautheit mit einem Gehörmodell. Proc. DAGA 2012, Darmstadt, 591–592. (2012)

[15] Sottek, R.: Modelle zur Signalverarbeitung im menschlichen Gehör. Dissertation, RWTH Aachen, (1993)

[16] Sottek, R., Lobato, T. and Becker, J.: Loudness of sounds with a subcritical bandwidth: improved prediction with the concept of tonal loudness. Proc. DAGA 2022, Stuttgart. (2022)

[17] Sottek, R., Lobato, T., Bender, M. and Becker, J.: Modelling the ISO 226: 2023 equal-loudness-level contours by standardized loudness methods. Proc. of the 10th Forum Acusticum. (2023)