ecma INTERNATIONAL

Standard ECMA-418-2

2nd Edition / December 2022

Psychoacoustic metrics for ITT equipment — Part 2 (models based on human perception)

Rue du Rhône 114 CH-1204 Geneva T: +41 22 849 6000 F: +41 22 849 6001

Contents

Page

Introduction

ECMA-418-2 specifies methods for identifying perceptually prominent components in airborne noise emitted by information technology and telecommunications (ITT) equipment using models of human perception. The content was originally published in ECMA-74 17th edition “Measurement of Airborne Noise emitted by Information Technology and Telecommunications Equipment”. Psychoacoustic content of ECMA-74 was moved to ECMA-418 Parts 1 and Part 2 to distinguish and separate it from the legacy prescriptions of microphone position, equipment operation, and sound level processing, which remain in ECMA-74.

ECMA-418 Parts 1 and 2 are psychoacoustic standards and as such prescribe methods that represent the perception of noise emitted by ITT equipment. Sound signals recorded by the procedures of ECMA-74 are analysed using the psychoacoustic methods of ECMA-418 Parts 1 and 2. While intended for ITT equipment, the methods may be useful for other applications as well.

The psychoacoustic methods in this standard, ECMA-418 Part 2 are based on a human hearing model of Sottek that expresses specific loudness, which describes level- and frequency-dependent masking and threshold of hearing. The model approximates the well-established Zwicker specific loudness method, but was extended by using a modified Bark scale covering the entire audible frequency range and an improved nonlinear matching of loudness at higher levels, which leads to a significant improvement of the prediction quality for several loudness matching experiments using synthetic and technical sounds.

Additional models described in this standard use the specific loudness to express the strength of perceived tonality and roughness. The models of this standard, Part 2, are more intricate than those of Part 1, which considers sound pressure in narrow and critical bands and hearing threshold.

The first edition of ECMA-418-2 was issued in December 2020.

For the 2nd edition, there were several updates as follows:

- The hearing model, tonality, and roughness procedures of Clauses 5, 6, and 7 were refined, and the descriptions of these procedures improved to assist implementation.
- In Clause 5, a figure showing auditory filter bank response of the hearing model of Sottek was added to assist implementation.
- An entropy weighted roughness based on modulation rate random was added to Clause 7.1 for applications in which measured rotational speed is available.
- Clause 8 was added to describe loudness of sounds with subcritical or larger bandwidths.

ECMA-418 series consists of the following parts, under the general title “Psychoacoustic metrics for ITT equipment”:

- Part 1 (prominent discrete tones)
- Part 2 (models based on human perception)

This Ecma Standard was developed by Technical Committee 26 and was adopted by the General Assembly of December 2022.

Psychoacoustic metrics for IT equipment − Part 2 (models based on human perception)

1 Scope

This standard describes the hearing model and psychoacoustic metrics dependent on the hearing model. The input to the hearing model are sound signals recorded using the procedures of ECMA-74. The hearing model expresses specific loudness [1]. Psychoacoustic models use the specific loudness to express the strength of any tonalities or roughness in the sound generated by Information Technology and Telecommunications (ITT) equipment. While developed for ITT equipment, the psychoacoustic methods of this standard may be relevant to other applications like automobiles, consumer appliances, etc.

The tonality metric of this standard uses the auto-correlation function to describe causes of perceived tonality such as individual or multiple steady or time-varying discrete tones, individual or multiple spectrally elevated bands or slopes of noise, and combinations of these phenomena. A similar approach was published in 1998 to determine “pitch salience” [2].

The roughness metric presented in this standard uses a spectrum of the sound signal envelope, refined by a quadratic fit estimator, to describe roughness arising from sound signal envelope variations within a critical band at modulation rates between 20 and around 300 Hz. For steady sounds, roughness perception peaks at modulation rates of 70 Hz.

The loudness metric of this standard uses a nonlinear combination of tonal and noise loudness calculated as intermediate results of the tonality algorithm to achieve a very good match of perceived loudness, especially for sounds with a subcritical bandwidth (sounds containing tonal and noise components).

2 Conformance

Measurements are in conformity with this Standard if they meet the following requirements:

a) The measurements are taken in conformity with the Standard ECMA-74.
b) The measurements are carried out with a sampling rate of 48 kHz or they are resampled to a sampling rate of 48 kHz if they were originally taken with a different sampling rate.
c) For the determination of prominent tonalities, the method specified in Clause 6 is used.
d) For the determination of prominent roughness, the method specified in Clause 7 is used.

3 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ECMA-74, *Measurement of Airborne Noise emitted by Information Technology and Telecommunications Equipment*, 19th edition (December 2021)

ISO 226:2003, *Acoustics — Normal equal-loudness-level contours*

4 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

NOTE If a definition is identical to that in another standard, that standard and definition number is given in brackets.

4.1
loudness
N
perceived magnitude of a sound, which depends on the acoustic properties of the sound and the specific listening conditions, as estimated by that the average human listener with normal hearing.

NOTE 1 Loudness is expressed in sones.

NOTE 2 Loudness depends primarily upon the sound pressure level, although it also depends upon the frequency, bandwidth, and duration of the sound.

NOTE 3 A sound that is twice as loud as another sound is characterized by doubling the number of sones.

NOTE 4 Adapted from ISO 532-1, 3.18

4.2
specific loudness
N'
perceived magnitude or volume of sound in a critical band.

NOTE 1 The unit of specific loudness is expressed in terms of sone per Bark.

4.3
equal loudness contour
the sound pressure level[1] for which the average human listener with normal hearing perceive constant loudness when presented with a single frequency (pure) tone.

NOTE 1 Equal loudness contour is parameterized by the sound pressure level and frequency of the presented tone. See ISO 226:2003.

4.4
threshold of hearing
level of a sound at which, under specified conditions, a person gives 50 % of correct detection responses on repeated trials.

[SOURCE: ISO 226 :2003, 3.7

4.5
critical band
filter within the human cochlea describing the frequency resolution of the auditory system with characteristics that are usually estimated from the results of masking experiments.

[SOURCE: ISO 532-1[3], 3.12]

4.6
critical bandwidth
bandwidth of a critical band.

NOTE 1 Each critical bandwidth has a width of one unit on the critical band rate scale.

[1] The definition of sound pressure level is given in the terms and definitions of ECMA-74.

4.7
critical band rate scale
transformation of the frequency scale, constructed so that an increase in frequency equal to one critical bandwidth leads to an increase of one unit on the critical band rate scale.

NOTE 1 Frequencies on the critical band rate scale are expressed in Bark.

NOTE 2 Adapted from ISO 532-1, 3.14

4.8
tonality
a characteristic of sound containing a single-frequency component or narrow-band components that emerge audibly from the total sound.

NOTE 1 Tonality can arise from individual or multiple steady or time-varying discrete tones, individual or multiple spectrally elevated bands or slopes of noise, and combinations of these phenomena.

4.9
envelope
the instantaneous amplitude of a signal.

NOTE 1 The instantaneous amplitude describes the low-frequent variations of the amplitude. It has a significantly lower frequency than the carrier frequency of the signal.

4.10
roughness
a characteristic of sound with the quality of being uneven yet steady.

NOTE 1 Roughness can arise if the envelope of a sound signal within a critical band has temporal variation.

4.11
modulation
fluctuation of the envelope of a signal over time.

NOTE 1 Modulation is expressed in terms of its strength (modulation index) and the speed at which it changes (modulation rate).

4.12
modulation rate
frequency of changes of the envelope of a signal.

NOTE 1 The modulation rate is expressed in Hertz.

NOTE 2 The word “rate” is used to avoid confusion with the sound frequency.

5 A hearing model approach to calculate psychoacoustic parameters

This clause describes a perception-model-based procedure for determining the specific loudness of a sound, the hearing model of Sottek. There are different loudness calculation procedures, such as the German standard DIN 45631/A1 [3] and the international standard ISO 532-1 [4] (both based on Zwicker's loudness model) as well as the Dynamic Loudness Model (by Chalupper and Fastl) [5], the Time Varying Loudness model (by Glasberg and Moore) [6], and the loudness calculation algorithm based on the hearing model of Sottek, allowing for the prediction of the perceived loudness of time-varying sounds in many cases (ISO 532-2[7] only applies to stationary sounds). However, previous studies of Rennies et al. [8], [9] showed that the predictions for some time-varying 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 in Reference [1] focusing on impulsive sounds. On the basis of these experiments, it was studied how far the hearing model approach to time-varying loudness according to Sottek can account for the specific signal properties of these sounds. It could be shown that the hearing model approach to time-varying loudness performs better than other existing loudness models: The hearing model, characterized especially by the application of an improved 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. In addition, the auditory filter bank used is based on an extended Bark scale covering the entire audible frequency range while matching the experimental results related to critical bandwidth better than other models. Further, the hearing model is able to predict the nonlinear behaviour with respect to just-noticeable amplitude differences and variations. [1]

The hearing model described in this clause transforms sound pressure to loudness, where the unit of loudness is $\mathrm{sone_{HMS}}$, where HMS stands for "according to the **H**earing **M**odel of **S**ottek" and denotes that the loudness differs from other definitions. The result of the hearing model can be used as the basis for further psychoacoustic analyses.

5.1 Psychoacoustic hearing model

5.1.1 Overview

Figure 1 displays the basic hearing model structure for calculating specific loudness as the basis for determining other psychoacoustic sensations. Subsequently, the different signal processing blocks of the hearing model are briefly explained.

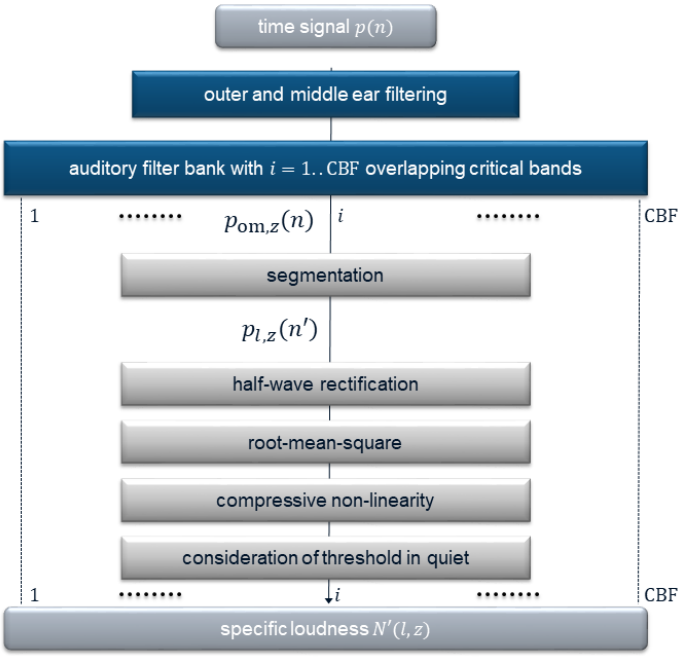

Figure 1 — Basic hearing model structure, including the auditory filter bank, where CBF is the number of critical band filters in the filter bank.

The input signal is a discrete time signal containing sound pressure values with a sampling rate of $r_s = 48$ kHz [2]. The common sampling rate $r_s = 48$ kHz is chosen to ensure that the entire audible frequency range is covered.

5.1.2 Pre-processing of input data

Initially, the first 5 ms of the input signal (corresponding to $n_{\text{fade in}} = 0{,}005 \cdot 48000 = 240$ samples) are multiplied with a trigonometric weighting function

$$w_{\text{fade in}}(n) = 0{,}5 - 0{,}5 \cdot \cos\left(\frac{\pi n}{n_{\text{fade in}}}\right) \quad (1)$$

with $n = 0, 1, \ldots, n_{\text{fade in}} - 1$ in order to reduce artifacts due to filter oscillations in case of signals starting with non-zero values.

Second, zero-padding on both ends of the signal shall be performed to facilitate later processing steps. The number of zeros at the end $n_{\text{zeros,end}}$ is calculated as:

$$n_{\text{zeros,end}} = n_{\text{new}} - n_{\text{samples}}, \quad (2)$$

where n_{samples} is the number of samples of the signal and n_{new} equals to:

$$n_{\text{new}} = s_{\text{h,max}} \cdot \left(\text{ceil}\left(\frac{n_{\text{samples}} + s_{\text{h,max}} + s_{\text{b,max}}}{s_{\text{h,max}}}\right) - 1\right), \quad (3)$$

where the $\text{ceil}(x)$ operator gives the smallest integer value higher than or equal to the number x. The band-dependent block size $s_b(z)$ and the hop size[3] $s_h(z)$ are defined in detail in Clause 5.1.5 and $s_{\text{b,max}}$ and $s_{\text{h,max}}$ are the largest band-dependent block size and hop size of all used filter stages, which are defined in Clause 6.2.2 for the tonality and in Clause 7.1.1 for the roughness. The number of zeros at the start $n_{\text{zeros,start}}$ shall be equal to $s_{\text{b,max}}$. The zero-padded sound pressure signal is named $p(n)$.

5.1.3 Outer and middle/inner ear filtering

5.1.3.1 Theory

The pre-processing consists of filtering the input signal $p(n)$ with transfer functions of the outer and of the middle/inner ear. The transfer function of the outer ear was modelled based on measured head related transfer functions (HRTFs). The transfer function of the middle/inner ear was chosen such that the filtering together with the loudness threshold $\text{LTQ}(z)$ (as explained in Clause 5.1.9) leads to a loudness estimation emulating the equal-loudness contours from 20 to 90 phon (with a step size of 10 phon) and the lower threshold of hearing.[4] The middle/inner ear filter is optimized on the equal-loudness contours of ISO 226:2003.

---

[2] If the input data is sampled at a different sampling rate than 48 kHz, a resampling to 48 kHz needs to be performed.

[3] The hop size is the time shift to the next calculation block, smaller than block size if overlapping is used. It is related to the percent overlap ov by $s_h(z) = s_b(z) \cdot (100 - \text{ov})/100$.

[4] In Zwicker's loudness model [3] the influence of the outer and middle ear transfer functions is considered by the ear's transmission characteristic a_0.

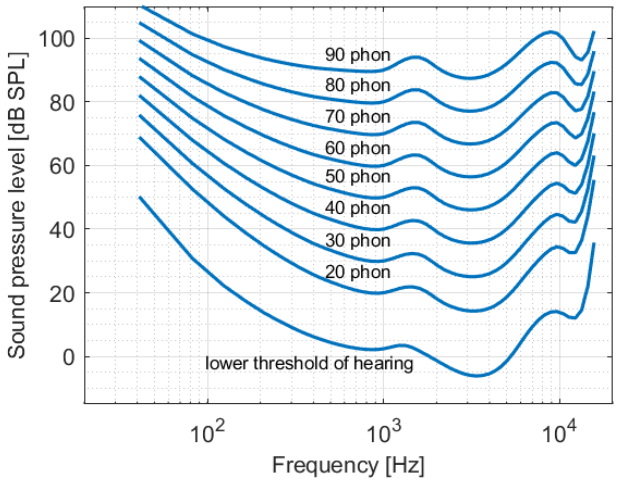

Figure 2 — Equal loudness contours (ISO 226:2003) used as target for the filter transfer function

The lower threshold of hearing is also taken from ISO 226:2003. This corresponds also to the data of the lower threshold of hearing published in ISO 389-7[10]. The target equal-loudness contours are illustrated in Figure 2. An evaluation of the hearing model showing the emulated equal-loudness contours is given in Annex A.

5.1.3.2 Implementation

The transfer function of the resulting filter is shown in Figure 3. The overall filter is composed of a filter modelling the influence of the outer ear and a filter modelling the influence of the middle/inner ear. Those filters are also shown in Figure 3.

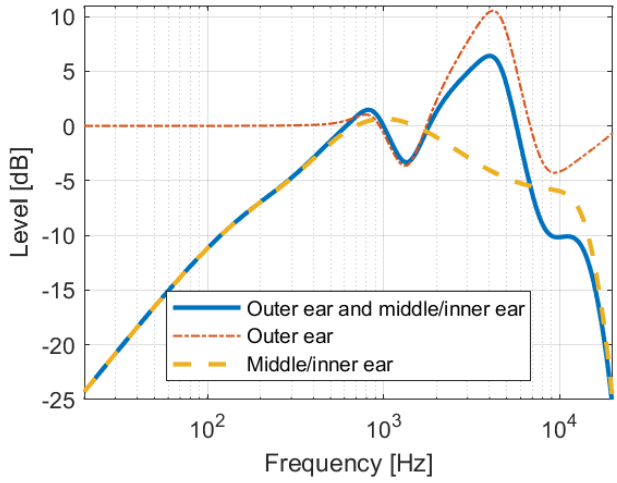

Figure 3 — Transfer function of the outer and middle/inner ear filter

For numerical reasons, it is recommended to implement this high-order filter as $K = 8$ serially-cascaded second-order filters $H_k(f)$. The filter function $H(f)$ is then defined as

$$H(f) = \prod_{k=1}^{K} H_k(f). \tag{4}$$

Each second-order filter $H_k(f)$ can be implemented using the recursive Formula (5)

$$y(n) = \sum_{m=0}^{2} b_{mk} x(n-m) - \sum_{m=1}^{2} a_{mk} y(n-m) \tag{5}$$

with input $x(n)$ and output $y(n)$. The corresponding filter coefficients are given in Table 1. The first five filters describe the influence of the outer ear, the last three describe the influence of the middle/inner ear. The filtering results in a filtered signal $p_{\mathrm{om}}(n)$.

Table 1 — Filter coefficients of outer and middle/inner ear filter

	Filter coefficients				
k	b_{0k}	b_{1k}	b_{2k}	a_{1k}	a_{2k}
1	1,015896	-1,925299	0,922118	-1,925299	0,938014
2	0,958943	-1,806088	0,876439	-1,806088	0,835382
3	0,961372	-1,763632	0,821788	-1,763632	0,783160
4	2,225804	-1,434650	-0,498204	-1,434650	0,727599
5	0,471735	-0,366092	0,244145	-0,366092	-0,284120
6	0,115267	0,000000	-0,115267	-1,796003	0,805838
7	0,988029	-1,912434	0,926132	-1,912434	0,914161
8	1,952238	0,162320	-0,667994	0,162320	0,284244

5.1.4 Auditory filtering bank

5.1.4.1 Theory

An auditory filter bank consisting of overlapping asymmetric filters 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 shape of the auditory filters matches the gammatone filters [11]. The amplitude is chosen such that the filter has a gain of 0 dB at the centre frequency $F(z)$, with z denoting the critical band rate scale. This 0 dB gain varies slightly for the first critical bands due to influence of the negative frequencies as seen in Figure 4. The critical bandwidth $\Delta f(z)$ is chosen such that it corresponds to the equivalent rectangular bandwidth (implementation details are given in Formulae (9) and (10)). 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. The impulse responses of the auditory filters are chosen as modulated low-pass filters (j is the imaginary unit):

$$h_z(t) = 2 \cdot \mathrm{Re}\left(h_{\mathrm{LP},z}(t) \cdot \exp(j2\pi F(z)t) \right) = 2 \cdot h_{\mathrm{LP},z}(t) \cdot \cos(2\pi F(z)t)\,. \tag{6}$$

The filters are calculated using the low-pass function

$$h_{\mathrm{LP},z}(t) = \varepsilon(t) \cdot \frac{1}{(k-1)!} \cdot \frac{1}{\tau(z)} \cdot \left(\frac{t}{\tau(z)}\right)^{k-1} \exp\left(-\frac{t}{\tau(z)}\right) \tag{7}$$

where k is the filter order[5], $\varepsilon(t)$ is the unit step function and the exclamation mark denotes the factorial operation. $\tau(z)$ is a time constant, related to $\Delta f(z)$ by

$$\tau(z) = \frac{1}{2^{2k-1}} \cdot \binom{2k-2}{k-1} \cdot \frac{1}{\Delta f(z)}. \tag{8}$$

The centre frequencies $F(z)$ and corresponding bandwidths $\Delta f(z)$ of the filter bank are calculated as

$$F(z) = \frac{\Delta f(f=0)}{c} \sinh(cz) \tag{9}$$

[5] Filter order $k = 5$ is used.

and

$$\Delta f(z) \quad = \quad \sqrt{\left(\Delta f(f=0)\right)^2 + (cF(z))^2}, \tag{10}$$

with z denoting the critical band rate scale.

Values for z are chosen from 0,5 to 26,5 with a step size of $\Delta z = 0{,}5$. $\Delta f(f = 0) = 81{,}9289$ Hz and $c = 0{,}1618$. These functions and settings lead to a better matching to the Bark table by Zwicker [12] than other existing formulae, as documented in detail in Reference [1]. The unit of the critical band rate scale of this auditory filter bank is $\mathrm{Bark}_{\mathrm{HMS}}$, where HMS stands for "according to the **H**earing **M**odel of **S**ottek" and denotes that the critical bands differ from other definitions.

As discrete approximation of the low-pass filter,

$$h_{\mathrm{LP},z}(n) \quad = \quad \varepsilon(n) \cdot \frac{(1-d)^k}{\sum_{i=1}^{k-1} e_i d^i} n^{k-1} d^n, \tag{11}$$

with time index n and $d = \exp\left(-\frac{1}{r_s \tau(z)}\right)$ is used[6]; e_i depends on the filter order k and is given below for a specific value of k. The band-pass filtering using $h_z(t)$ can be implemented using the discrete approximation of the band-pass filter

$$h_z(n) \quad = 2 \cdot \mathrm{Re}\left(h_{\mathrm{LP},z}(n) \cdot \exp\left(\frac{j2\pi F(z) n}{f_s}\right)\right) = 2 \cdot h_{\mathrm{LP},z}(n) \cdot \cos\left(\frac{2\pi F(z) n}{f_s}\right). \tag{12}$$

5.1.4.2 Implementation

In the following, instructions for the implementation of the auditory filters are given: Digital filtering can be implemented using the recursive Formula (13):

$$y(n) \quad = \quad \sum_{m=0}^{k-1} b_m x(n-m) - \sum_{m=1}^{k} a_m y(n-m)\,. \tag{13}$$

For the discrete low-pass filter $h_{\mathrm{LP},z}(n)$ as described in Formula (11), the real-valued filter coefficients are

$$a_m \quad = \quad (-d)^m \binom{k}{m}, \tag{14}$$

and

$$b_m \quad = \quad \frac{(1-d)^k}{\sum_{i=1}^{k-1} e_i d^i} d^m e_m. \tag{15}$$

With a used filter order of $k = 5$ the coefficients e_i in Formula (11) and in Formula (15) are given as $e_0 = 0, e_1 = 1, e_2 = 11, e_3 = 11,$ and $e_4 = 1$. As explained above, $d = \exp\left(-\frac{1}{r_s \tau(z)}\right)$ with $\tau(z)$ as defined in Formula (8).

[6] $r_s = 48$ kHz is the sampling rate.

The coefficients a_m and b_m can be used for the implementation of the discrete approximation of the low-pass function given in Formulae (7) and (11). However, to obtain the discrete approximation of the band-pass filter in Formula (12), the filter coefficients of the low-pass filter shall be modified to:

$$a'_m = a_m \exp\left(\frac{j2\pi F(z)m}{r_s}\right) \tag{16}$$

and

$$b'_m = b_m \exp\left(\frac{j2\pi F(z)m}{r_s}\right), \tag{17}$$

with a sampling rate of $r_s = 48$ kHz. Using these modified filter coefficients in the recursive Formula (13) results in a discrete implementation of the auditory filters. The filter results in a complex-valued band-pass signal with a single-sided spectrum. Two times the even part of the spectrum of this signal corresponds to the real-valued band-pass signal. Thus, the real-valued band-pass signal can be determined as the double real part of the complex result.

Figure 4 shows the magnitude of the transfer functions of the auditory filter bank, calculated by filtering a digital Dirac pulse (sampling rate: 48000 Hz, duration 1 s) using the filter coefficients[7] defined in Formulae (16) and (17) with a subsequent Fourier transform on the real-value band-pass signal.

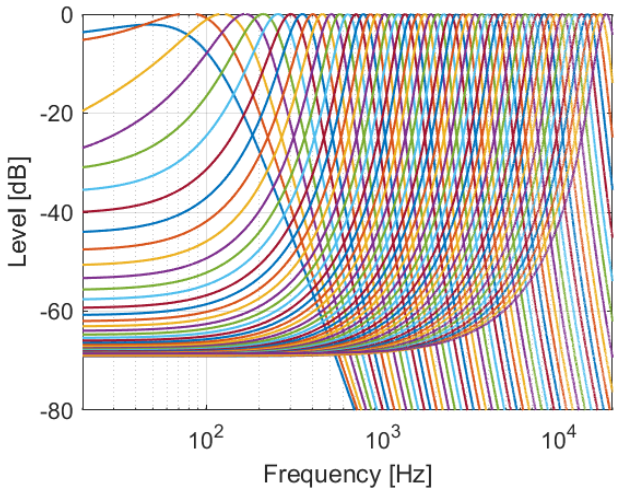

Figure 4 — Magnitude of the transfer functions of the auditory filter bank

The auditory filter bank results in $\mathrm{CBF} = 53$ band-pass signals $p_{\mathrm{om},z}(n)$ centred around the critical band rate scale values z ranging from 0,5 to 26,5, thus leading to an extension of the Bark scale for frequencies of the entire audibility range up to approximately 20 kHz using 53 critical band filters with an overlap of 50%.

5.1.5 Segmentation

For further processing, segmentation into blocks needs to be performed and blockwise root-mean-square (RMS) values need to be calculated. For the segmentation, the band-dependent block size $s_b(z)$ and the hop size $s_h(z)$ can be chosen depending on the application. Values for $s_b(z)$ and $s_h(z)$ for the calculation of the psychoacoustic tonality are given in Clause 6.2.2 and for the calculation of the psychoacoustic roughness in Clause 7.1.1.

[7] Filtering shall be performed with double precision.

The segmentation can be described as:

$$p_{l,z}(n') \quad = \quad p_{\mathrm{om},z}(l \cdot s_{\mathrm{h}}(z) + i_{\mathrm{start}}(z) + n') \tag{18}$$

with $0 \leq n' \leq s_{\mathrm{b}}(z) - 1$, where the time index l describes the block number of each block, starting with $l = 0$ (corresponding to a time of 0 ms). $i_{\mathrm{start}}(z)$ is an index offset that guarantees that the first block of all stages corresponds to the same time reference. It is defined as:

$$i_{\mathrm{start}}(z) \quad = s_{\mathrm{b}}(0.5) - s_{\mathrm{b}}(z). \tag{19}$$

Thus, each block $p_{l,z}(n')$ ranges from $n = l \cdot s_{\mathrm{h}}(z) + i_{\mathrm{start}}(z)$ to $n = l \cdot s_{\mathrm{h}}(z) + i_{\mathrm{start}}(z) + s_{\mathrm{b}}(z) - 1$. The last value of l, $l_{\mathrm{last}}(z)$, is dependend on the filter band and the value n_{new} defined in Formula (3):

$$l_{\mathrm{last}}(z) \quad = \quad \mathrm{ceil}\left(\frac{n_{\mathrm{new}} + s_{\mathrm{h}}(z)}{s_{\mathrm{h}}(z)}\right) - 1\,. \tag{20}$$

5.1.6 Rectification

Subsequent half-wave rectification accounts for the fact that the auditory nerves fire only when the basilar membrane vibrates in a specific direction [13]. The resulting band-pass signals are calculated as:

$$p_{\mathrm{rect},l,z}(n') = \begin{cases} p_{l,z}(n'), & p_{l,z}(n') > 0 \\ 0, & p_{l,z}(n') \leq 0 \end{cases}. \tag{21}$$

5.1.7 Calculation of root-mean-square values

With the segmented and rectified blocks $p_{\mathrm{rect},l,z}(n')$, the RMS-values are calculated for each block as:

$$\tilde{p}(l,z) \quad = \quad \sqrt{\frac{2}{s_{\mathrm{b}}(z)} \sum_{n'=0}^{s_{\mathrm{b}}(z)-1} {p_{\mathrm{rect},l,z}}^2(n')}\,, \tag{22}$$

The factor of 2 is necessary to compensate for the signal energy which was lost due to the half-wave rectification. The dependency on the time index l is dropped in the following, since the further processing steps are applied to each time block in the same way.

5.1.8 Nonlinearity to transform sound pressure into specific loudness

The compressive 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 tonality, roughness or fluctuation. Such a nonlinearity function has proven applicable to predict many phenomena like ratio loudness, just-noticeable amplitude differences and modulation thresholds as well as the level dependence of roughness.

The nonlinearity between specific loudness and sound pressure was reconsidered in the hearing model according to results of many listening tests [14]. Further improvements for higher levels above approximately 80 dB were achieved by introducing a nonlinearity function according to Formula (23):

$$A'(\tilde{p}) \quad = \quad c_{\mathrm{N}} \cdot \left(\frac{\tilde{p}}{\tilde{p}_0}\right) \cdot \prod_{i=1}^{M} \left(1 + \left(\frac{\tilde{p}}{\tilde{p}_{t_i}}\right)^{\alpha}\right)^{\frac{v_i - v_{i-1}}{\alpha}} \tag{23}$$

with root-mean-square values of sound pressure $\tilde{p}$ and thresholds $\tilde{p}_{t_i}$ in Pa, $\tilde{p}_0 = 20$ µPa. The M thresholds $\tilde{p}_{t_i}$ can be derived from Table 2; α is set to 1,5; $c_{\mathrm{N}} = 0{,}0211668$ is a calibration factor with the

unit $\mathrm{sone_{HMS}/Bark_{HMS}}$[8] to assure that the total loudness of a sinusoid having a frequency of 1 kHz and a sound pressure level of 40 dB equals 1 $\mathrm{sone_{HMS}}$ (using the method described in Clause 8.1)[9]. The $M = 8$ exponents v_i as given in Table 2 were achieved by applying a nonlinear-optimization procedure in order to minimize the root-mean-square error between the results of the loudness matching experiment and the results of the model calculation. The initial exponent v_0 is set to 1.

Table 2 — $M = 8$ thresholds and exponents for the nonlinearity function for Formula (23)

i	1	2	3	4	5	6	7	8
$20\log_{10}(\tilde{p}_{t_i}/\tilde{p}_0)$ [dB]	15	25	35	45	55	65	75	85
v_i	0,6602	0,0864	0,6384	0,0328	0,4068	0,2082	0,3994	0,6434

The nonlinearity is applied to $\tilde{p}(z)$ in each band z. The resulting variable

$$\tilde{N}'(z) \quad = \quad A'\big(\tilde{p}(z)\big) \tag{24}$$

can be interpreted as the specific loudness of the signal without consideration of the threshold in quiet.

The function according to Formula (23) results from an optimization procedure to fit the experimental data with the lowest root-mean-square error [14]. It has a steep slope at high levels, which agrees with results of experiments from Buus et al. [15] and Epstein et al. [16]

5.1.9 Consideration of threshold in quiet

The specific loudness in each band z is zero if it is at or below a critical-band-dependent specific loudness threshold $\mathrm{LTQ}(z)$. The band-specific loudness threshold $\mathrm{LTQ}(z)$ is given for each used band number z from 0,5 to 26,5 in Table 3. Figure 5 shows the loudness threshold $\mathrm{LTQ}(z)$ in dependency of the center frequency of the bands.

Table 3 — Specific loudness threshold $\mathrm{LTQ}(z)$ for each used value of z

z	$\mathrm{LTQ}(z)$	z	$\mathrm{LTQ}(z)$	z	$\mathrm{LTQ}(z)$	z	$\mathrm{LTQ}(z)$	z	$\mathrm{LTQ}(z)$
0,5	0,3310	6,0	0,0151	11,5	0,0071	17,0	0,0122	22,5	0,0202
1,0	0,1625	6,5	0,0131	12,0	0,0072	17,5	0,0138	23,0	0,0217
1,5	0,1051	7,0	0,0115	12,5	0,0073	18,0	0,0157	23,5	0,0237
2,0	0,0757	7,5	0,0103	13,0	0,0074	18,5	0,0172	24,0	0,0263
2,5	0,0576	8,0	0,0093	13,5	0,0076	19,0	0,0180	24,5	0,0296
3,0	0,0453	8,5	0,0086	14,0	0,0079	19,5	0,0180	25,0	0,0339
3,5	0,0365	9,0	0,0081	14,5	0,0082	20,0	0,0177	25,5	0,0398
4,0	0,0298	9,5	0,0077	15,0	0,0086	20,5	0,0176	26,0	0,0485
4,5	0,0247	10,0	0,0074	15,5	0,0092	21,0	0,0177	26,5	0,0622
5,0	0,0207	10,5	0,0073	16,0	0,0100	21,5	0,0182		
5,5	0,0176	11,0	0,0072	16,5	0,0109	22,0	0,0190		

[8] HMS stands for “according to the **H**earing **M**odel of **S**ottek” and denotes that the calculated loudness and the critical bands differ from other definitions.

[9] The calibration factor c_{N} can be adjusted within a tolerance of 0,25 % to account for the effects of different implementations.

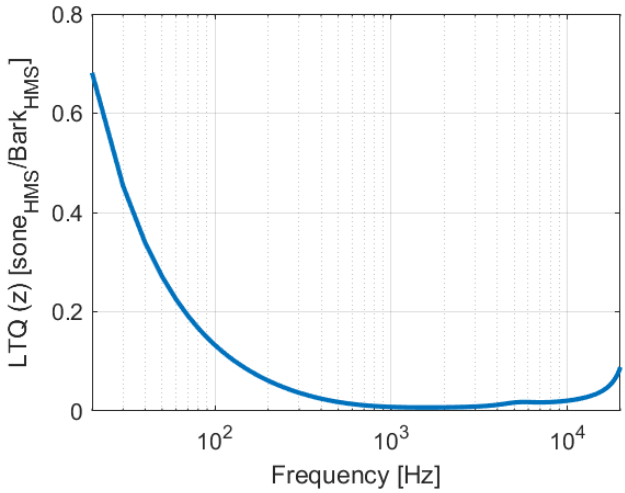

Figure 5 — Specific loudness threshold $\mathbf{LTQ}(z)$

The lower threshold of hearing is applied by subtraction and a limiter:

$$N'_{\text{basis}}(z) \quad = \quad \begin{cases} \widetilde{N}'(z) - \text{LTQ}(z), & \widetilde{N}'(z) \geq \text{LTQ}(z) \\ 0 & \widetilde{N}'(z) < \text{LTQ}(z) \end{cases}. \tag{25}$$

The result $N'_{\text{basis}}(z)$ is the specific basis loudness of the signal. The specific basis loudness can be used as basis for other psychoacoustic parameters such as tonality (see Clause 6) and roughness (see Clause 7).

A signal is considered to be audible when its total loudness value exceeds 0,01 sone_{HMS}, where total basis loudness is calculated summing all specific basis loudness values, using $\Delta z = 0{,}5$ as

$$N_{\text{basis}} \quad = \quad \sum_{i=1}^{\text{CBF}} N'_{\text{basis}}\left(\frac{i}{2}\right) \cdot \Delta z\,. \tag{26}$$

Consideration of both total and specific basis loudness has the benefit of allowing loudness summation of sounds consisting of multiple components near threshold.

Recent investigations showed that existing loudness procedures underestimate the loudness of tonal signals [17]. Clause 8.1 describes a new loudness algorithm based on a nonlinear weighting of the partial loudness of tonal and non-tonal components derived in Clause 6.2.

6 Identification and evaluation of prominent tonalities using a psychoacoustic tonality calculation method

This clause describes a perception-model-based procedure for determining whether or not noise emissions contain prominent tonalities, and if present, their strengths: the psychoacoustic tonality calculation method. A similar approach was published in 1998 for the determination of “pitch salience” [2]. The calculation is based on the specific basis loudness as described in Clause 5.

Prominent perceived tonalities arise from a variety of causes including but not limited to prominent discrete tones: discrete tones, non-pure tones, narrow elevated noise bands, combinations of tones and narrow elevated noise bands, band-edges of various slopes terminating elevated noise bands of various bandwidths, and combinations of these. This clause defines a procedure for identifying and ranking tonalities from any causes.

6.1 Determination of tonality

6.1.1 Tonalities and their relationships to the threshold of hearing

Discrete tones or other tonalities should only be classified as *prominent* if they are, in fact, *audible* in the noise emissions of the equipment under test. For the tonality calculation methods as described in ECMA-418 – Part 1: Dominant discrete tones, a pre-calculation screening test is recommended concerning audibility of the tonality. From calibrated acoustical measurement time-data, this step is not required with the psychoacoustic tonality calculation method regardless of proximity to the threshold of hearing because the method inherently considers the threshold of hearing and the psychoacoustic loudness of tonal and non-tonal components.

6.1.2 Multiple tones in a critical band, and time-variation of tonality due to their interaction

The noise emitted by a machine may contain multiple tones or narrowband tonalities, several of which may fall within a single critical band. Besides the likelihood of increased overall tonality strength due to a plurality of tones within one critical bandwidth, there is a strong likelihood of beating interference between or among the plural tonalities causing time structure (amplitude modulation): periodic additions and cancellations affecting the strength of the perceived tonality within that critical band. In this case the sound is often perceived as “rough”, leading to the psychoacoustics sensation of “roughness”. A method for the identification of prominent roughness is described in Clause 7.

6.2 Psychoacoustic tonality calculation method

6.2.1 Overview

Tonality perceptions arising from spectrally-elevated noise bands of various widths and slopes and from non-pure tones as well as from discrete (pure) tones, and from combinations of these, can be mis-measured or escape measure in “hybrid” sound pressure based tools and tools sensitive only to discrete tones. To address such issues, a new psychoacoustically-based tonality calculation method based on the hearing model in Clause 5[18] was developed. The applicability of the model was investigated for technical sounds and compared to established methods of tonality calculation [19], [20], [21]. The method automatically considers the threshold of hearing because the hearing threshold is built into the hearing model [21].

Recent research results show a strong correlation between tonality perception and the partial loudness of tonal sound components [22], [23], [24]. Therefore, the new hearing model approach to tonality on the basis of the perceived loudness of tonal content has been developed. The new model evaluates the nonlinear and time-dependent specific loudness of both tonal and broadband components, which are separated using the autocorrelation function. This model has been validated by many sound situations and listening tests [19].

In early publications, Licklider assumed that human pitch perception is based on both spectral and temporal cues [25]. According to Licklider, the neuronal processing in human hearing applies a running autocorrelation analysis of the critical band signals. Under this assumption, psychoacoustic tonality phenomena like difference-tone perception or the missing-fundamental phenomenon (”virtual pitch”) can be explained.

This work inspired the idea to use the sliding autocorrelation function as a processing block in the hearing model for the calculation of roughness and fluctuation strength [20], [26], [27] and later for other psychoacoustic quantities like tonality [19] and loudness [1]. The psychoacoustic tonality calculation is based on scaled ACFs $\varphi_z{}'(m)$ (see Clause 6.2.2, with z denoting the critical band rate scale and m denoting the lag), which are calculated using the specific basis loudness $N'_{\mathrm{basis}}(z)$ (see Formula (25)) and the $\mathrm{CBF} = 53$ rectified band-pass signals $p_z(n)$ (see Clause 5.1.6) as described in Clause 5. An evaluation of the psychoacoustic tonality method, including application examples, is given in Annex B.

The further processing for tonality calculation is performed similarly as published in References [19], [20], and [21] as shown in Figure 6 and described in detail as follows:

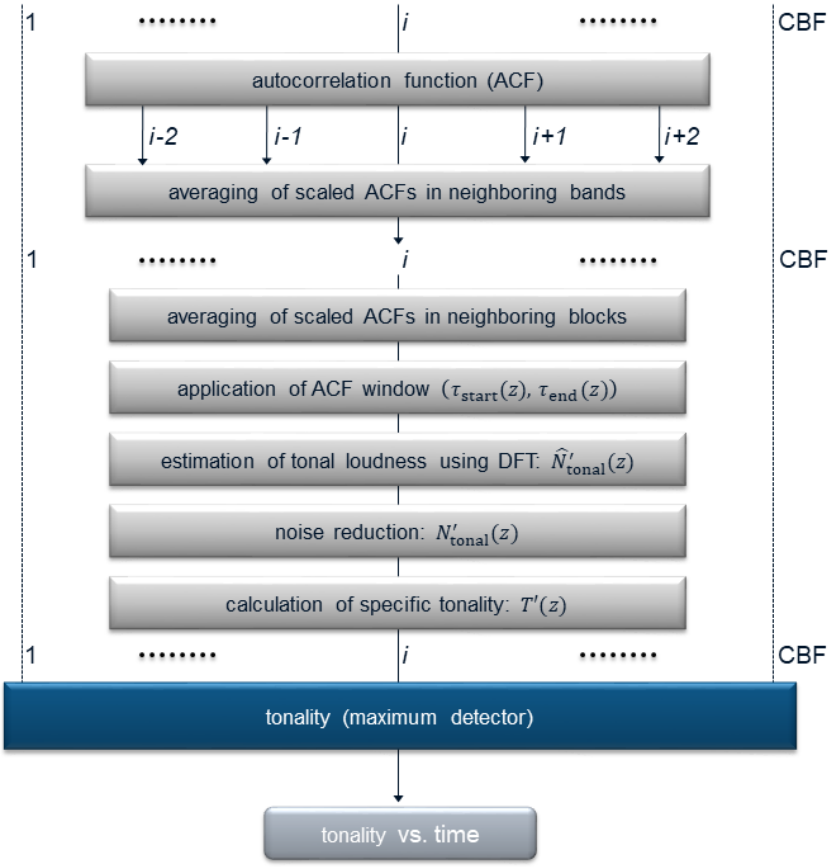

Figure 6 — Calculation of tonality based on the scaled ACFs as described in Reference [19], but with frequency-dependent analysis window borders

6.2.2 Autocorrelation function

Recently, it was proposed to use the autocorrelation function of the band-pass signals to separate tonal content from noise [1]. The autocorrelation function of white Gaussian noise is characterized by a Dirac impulse. Any broadband noise signal has at least a non-periodic autocorrelation function with high values at low lags, whereas the autocorrelation function (ACF) of periodic signals shows also a periodic structure [28]. Thus, the loudness of the tonal component can be estimated by analyzing the ACF at a certain range with respect to the lag m, and also the loudness of the remaining (noisy) part.The calculation of the sliding ACF is time-consuming. Therefore, the sliding ACF is calculated block-wise using the discrete Fourier Transform (DFT) to shorten computing time. An overlap of 75% is used for neighbouring blocks. There is a low-pass effect due to averaging over the block length. The ACF is performed on the same rectified blocks $p_{\mathrm{rect},l,z}(n')$ (see Formula (21)) of the overlapping critical band signals, from which the root-mean-square values were calculated in Clause 5.1.7.

For slowly varying low-frequency band-pass signals, a greater block length $s_{\mathrm{b}}(z)$ is necessary than for higher-frequency bands. Thus, different block lengths are used, depending on the frequency band. The block length is

chosen corresponding to the bandwidth $\Delta f(z)$ of each frequency band as described in Formula (10). The given values for the block size $s_\mathrm{b}(z)$ and the hop size $s_\mathrm{h}(z)$ also need to be used for the segmentation for the loudness calculation (see Clause 5.1.5).

Table 4 — Block length $s_\mathrm{b}(z)$ and hop size $s_\mathrm{h}(z)$ for the calculation of the autocorrelation function

$\Delta f(z)$	$0-$ 85 Hz	$85-170$ Hz	$170-340$ Hz	>340 Hz
z	$0{,}5-1{,}5$	$2-8$	$8{,}5-12{,}5$	≥ 13
$s_\mathrm{b}(z)$	8192	4096	2048	1024
$s_\mathrm{h}(z)$	2048	1024	512	256

For each block of length $s_\mathrm{b}(z)$, an unscaled autocorrelation function $\varphi_{l,z}(m)$ is calculated in two steps: first a $2s_\mathrm{b}$-point DFT[10] of $p_{\mathrm{rect},l,z}(n')$ is performed by zero padding, where $s_\mathrm{b}(z)$ is the block size given in Table 4, with a subsequent calculation of the squared magnitude:

$$P_{\mathrm{rect},l,z}(k) \quad = \quad \left|\mathrm{DFT}_{2s_\mathrm{b}}\left(p_{\mathrm{rect},l,z}(n')\right)\right|^2, \qquad 0 \leq k < 2s_\mathrm{b}(z)\,, \tag{27}$$

and second the **I**nverse **D**iscrete **F**ourier **T**ransform (IDFT[11]) of $P_{\mathrm{rect},l,z}(k)$ is calculated[12]:

$$\varphi_{\mathrm{unscaled},l,z}(m) \quad = \quad \mathrm{IDFT}_{2s_\mathrm{b}}\left(P_{\mathrm{rect},l,z}(k)\right), \qquad 0 \leq m < 2s_\mathrm{b}(z). \tag{28}$$

The next step is to compute a new estimate of an unbiased normalized autocorrelation function that compensates for lower overlaps at higher lag m (windowed, only values for $0 \leq m < \frac{3}{4}s_\mathrm{b}(z)$ needed):[13]

$$\varphi_{l,z}(m) = \begin{cases} \dfrac{\varphi_{\mathrm{unscaled},l,z}(m)}{\sqrt{\sum_{n'=0}^{s_\mathrm{b}(z)-m-1} {p_{\mathrm{rect},l,z}}^2(n') \cdot \sum_{n'=0}^{s_\mathrm{b}(z)-m-1} {p_{\mathrm{rect},l,z}}^2(n'+m)} + \varepsilon}, & 0 \leq m < \dfrac{3}{4}s_\mathrm{b}(z) \\ 0, & \dfrac{3}{4}s_\mathrm{b}(z) \leq m < 2s_\mathrm{b}(z) \end{cases}, \tag{29}$$

---

[10] The N-point DFT is defined as $X(k) = \mathrm{DFT}_N\big(x(n)\big) = \sum_{n=0}^{N-1} x(n) \cdot \mathrm{e}^{-j2\pi kn/N}$.

[11] The K-point IDFT is defined as $x(n) = \mathrm{IDFT}_K\big(X(k)\big) = \frac{1}{K}\sum_{k=0}^{K-1} X(k) \cdot \mathrm{e}^{+j2\pi kn/K}$.

[12] The presented calculations use two-sided spectra. This must be considered in an implementation since some signal processing libraries also use symmetry properties in their function calls to speed up the calculation and thus expect adjusted call parameters.

[13] A common problem in estimating a blockwise autocorrelation function is the decreasing overlap of the blocks with increasing lag m. The unscaled autocorrelation

$$\varphi_z(m) \quad = \quad \sum_{n'=0}^{s_\mathrm{b}(z)-m-1} p_z(n')p_z(n'+m)$$

does not consider this problem and thus leads to decreasing values for higher lag values, even if the signal is perfectly periodic. The commonly used approach for the unbiased autocorrelation, which aims to compensate for this problem, is

$$\varphi_z(m) \quad = \quad \frac{1}{s_\mathrm{b}(z)-|m|} \sum_{n'=0}^{s_\mathrm{b}(z)-m-1} p_z(n')p_z(n'+m)\,.$$

However, this approach may lead to unwanted effects, since the result does not necessarily satisfy the condition $\varphi_z(m) \leq \varphi_z(0)$, which is an essential property of the ACF. The new approach for the unbiased autocorrelation solves this problem by considering the energies of the overlapping parts of the blocks [29]. A drawback of this approach is the overestimation of the ACF of noise signals for higher lag values, but these values are neglected in further processing.

where the additive constant $\varepsilon = 10^{-12}$ prevents division by zero[14]. The dependency on the time index l is dropped in the following, since the further processing steps are applied to each time block in the same way.

The autocorrelation function has to to be calculated with two different block lengths for some frequency bands to allow averaging over neighbouring bands in later processing steps, as explained in the following Clause 6.2.3.

The entire ACF is multiplied with the specific basis loudness of the signal[15]:

$$\varphi_z'(m) \quad = \quad N'_{\text{basis}}(z) \cdot \varphi_z(m), \tag{30}$$

resulting in scaled[16] ACFs $\varphi'_z(m)$ which can be used for further analysis of the tonality.

6.2.3 Averaging of ACFs

First, ACFs of neighbouring bands are averaged in order to reduce noise. Averaging is performed over $2NB + 1$ bands, i.e., each band is averaged with the neighbouring NB lower and NB higher frequency bands. The value NB is chosen depending on the block size as described in Table 5. Since averaging needs to be performed with identical block size, it needs to be ensured that the autocorrelation function of neighbouring bands is available in the same block size. Thus, for frequency bands close to block size changes, the autocorrelation function needs to be calculated with two different block sizes. If not enough neighbouring frequency bands exist (for the lower frequency bands), NB is reduced such that averaging is still performed symmetrically centred around the particular frequency band. An exception is made for the lowest frequency band, which is averaged only with the second-lowest frequency band. This is necessary, because a symmetric averaging is not possible because of the missing lower band. No averaging on the other hand results in high noise artefacts.

Table 5 — Number of bands to average NB depending on block size s_b

s_b	8192	4096	2048	1024
NB	2	2	1	0

In a next step, the ACFs are averaged over neighbouring blocks in time for further reduction of noise. This block averaging is performed only for the block sizes $s_\text{b} = 8192$ and $s_\text{b} = 4096$, in which case the ACF in a given block is averaged with the ACFs in the preceding and the subsequent blocks. The averaging is not performed for the first and the last block because there is no preceding respectively nor subsequent block.

The outcome of the two averaging steps is a modified, noise reduced scaled ACF $\bar{\varphi}_z'(m)$.

6.2.4 Application of ACF window

A lag window with frequency-dependent limits ($\tau_{\text{start}}(z)$ and $\tau_{\text{end}}(z)$) according to Formulae (31) and (32) is applied to the ACF $\bar{\varphi}_z'(m)$ to separate tonal from noisy content:

$$\tau_{\text{start}}(z) \quad = \quad \max\left(\frac{0{,}5}{\Delta f(z)}, \tau_{\min}\right), \tag{31}$$

$$\tau_{\text{end}}(z) \quad = \quad \max\left(\frac{4}{\Delta f(z)}, \tau_{\text{start}}(z) + 1\ \text{ms}\right). \tag{32}$$

---

[14] The additive constant $\varepsilon = 10^{-12}$ is used throughout the complete document to avoid division by zero in several formulae.

[15] $N'_{\text{basis}}(z)$ is the specific loudness calculated in Formula (25).

[16] The ACF is scaled such that $\varphi_z'(0)$ represents the specific loudness $N'(z)$.

Here $\Delta f(z)$ is the bandwidth of the critical band centred at z, $\tau_{\min}$ is 2 ms.

It can be shown that the autocorrelation function of a periodic signal is itself periodic [28]. In the case of a pure tone, the period of the ACF equals the period of the tone. Consequently, the signal energy of a pure tone can be identified at multiples of the signal period. For white Gaussian noise, the autocorrelation function is a Dirac impulse, weighted by the power spectral density of the noise [28]. In case of broadband white noise, the autocorrelation function converges towards a Dirac impulse.

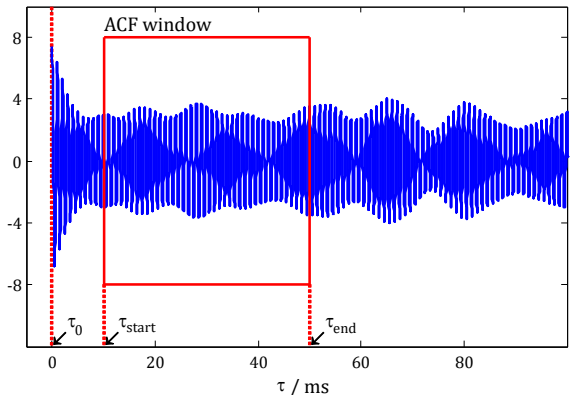

Figure 7 — Positioning of the ACF window for tonal content separation. This example shows the autocorrelation function of a tone in pink background noise

Figure 7 visualizes the placement of the ACF window for the autocorrelation function of a tone in pink background noise.[17]

From the calculated lag times, indices are calculated as

$$m_{\text{start}}(z) = \text{ceil}(\tau_{\text{start}}(z) \cdot r_s) - 1, \tag{33}$$

$$m_{\text{end}}(z) = \text{floor}(\tau_{\text{end}}(z) \cdot r_s) - 1, \tag{34}$$

where the $\text{ceil}(x)$ operator gives the smallest integer value higher than or equal to the number x and the $\text{floor}(x)$ operator gives the greatest integer value smaller than or equal to the number x. The window is applied by setting all elements of $\bar{\varphi}_z{}'(m')$ except the ones from index $m_{\text{start}}(z)$ to index $m_{\text{end}}(z)$ to zero and subtracting the mean of the windowed part of the ACF:

$$\varphi'_{z,\tau}(m) = \begin{cases} \bar{\varphi}_z{}'(m) - \dfrac{\sum_{m=m_{\text{start}}(z)}^{m_{\text{end}}(z)} \bar{\varphi}_z{}'(m)}{M}, & m_{\text{start}}(z) \le m \le m_{\text{end}}(z) \\ 0, & \text{else} \end{cases} \tag{35}$$

$M = m_{\text{end}}(z) - m_{\text{start}}(z) + 1$ is the number of samples in the window.

[17] The motivation for the limits given in Formulae (31) and (32) is as follows: In Figure 7, the energy distribution at small lags results from the noisy background and is disregarded by appropriately choosing the lower window border. Nevertheless, narrow-band noise also causes a perception of tonality when the bandwidth is comparatively small (i.e., few critical bands). This effect leads to a trade-off in the placement of the window borders: For a smaller bandwidth, the effect of the low-pass filtered noise on the ACF reaches higher lags than for a larger bandwidth. Thus, the window needs to be moved to higher lags for a lower bandwidth. On the other hand, higher lags are less reliable because they are calculated from a smaller number of samples. Therefore, the upper limit of the window should not be chosen too large.

6.2.5 Estimation of tonal loudness

The specific loudness of the tonal component is estimated by evaluating the spectrum of the ACF inside the lag window $\varphi'_{z,\tau}(m')$. A 16384-point DFT[18] of the M samples is performed by zero-padding, where the number 16384 is chosen as two times the largest block size $s_\mathrm{b}(z)$ given in Table 4:

$$\Phi'_{z,\tau}(k) \quad = \quad \mathrm{DFT}_{16384}\left(\varphi'_{z,\tau}(m')\right). \tag{36}$$

The maximum magnitude of the spectrum is searched, meaning, that the largest tonal content is extracted[19]:

$$\widehat{N}'_{\mathrm{tonal}}(z) \quad = \quad \begin{cases} 2\dfrac{\max\limits_k\left(\left|\Phi'_{z,\tau}(k)\right|\right)}{\frac{M}{2}}, & 2\dfrac{\max\limits_k\left(\left|\Phi'_{z,\tau}(k)\right|\right)}{\frac{M}{2}} \leq \bar{\varphi}_z{}'(0) \\ \bar{\varphi}_z{}'(0), & \text{else} \end{cases}. \tag{37}$$

$\widehat{N}'_{\mathrm{tonal}}(z)$ is a first estimation of the specific loudness of the tonal component. The frequency $f_{\mathrm{ton}}(z)$ of this component in the critical band centred around z can be estimated by first finding the DFT index $k_{\max}$ corresponding to the maximum of $\Phi'_{z,\tau}(k)$.

$$k_{\max}(z) \quad = \quad \arg\max_k\left(\Phi'_{z,\tau}(k)\right). \tag{38}$$

and calculating the corresponding frequency

$$f_{\mathrm{ton}}(z) \quad = \quad k_{\max}(z) \cdot \frac{r_\mathrm{s}}{16384}. \tag{39}$$

While this approach is capable of analysing tonalities with a rather high frequency resolution, it might underestimate tonal content when the corresponding frequency changes quickly inside of one block. This should be considered, even though the adaptive block size with smaller blocks for high frequencies aims at reducing this problem, since quickly varying frequencies usually occur at high frequencies.

6.2.6 Resampling to common time basis

For the further processing, the dependency of the time of each processed block becomes important. Thus, the time index l (which was dropped in Clause 6.2.2) needs to be considered. Since the results of different bands are in a different time basis at this stage of the processing due to a different block length, the bands with a higher block size are resampled to correspond to the time basis of the blocks calculated with the smallest block size of 1024. The resampling is done by linear interpolation. In Table 6, the interpolation factors i for each critical band z are given.

Table 6 — Interpolation factors for critical bands with different block size

z	$0{,}5-1{,}5$	$2-8$	$8{,}5-12{,}5$	≥ 13
$s_\mathrm{b}(z)$	8192	4096	2048	1024
i	8	4	2	1

For all time-dependent variables $(i-1)$ new samples are inserted between two given adjacent samples by simple linear interpolation.

---

[18] The N-point DFT is defined as $X(k) = \mathrm{DFT}_N\left(x(n)\right) = \sum_{n=0}^{N-1} x(n) \cdot \mathrm{e}^{-j2\pi kn/N}$.

[19] The normalization by $\frac{M}{2}$ is necessary to calculate the energy of the windowed ACF from the DFT result. The scaling factor 2 is necessary because of the half-wave rectified signal.

The final time index l is the one corresponding to the original time index of the smallest block size. In the following, $\widehat{N}'_{\mathrm{tonal}}(l,z)$ denotes the estimation of the specific loudness of the tonal component in the critical band centred around z at time index l. The sampling rate of these estimations is $r_{\mathrm{sd}} = \frac{r_{\mathrm{s}}}{s_{\mathrm{h,min}}} = \frac{48000\ \mathrm{Hz}}{256} = 187{,}5\ \mathrm{Hz}$. Here, the results belonging to the zero-padding done at the start of the processing need to be removed. Thus the last evaluated block shall be:

$$l_{\mathrm{end}} \quad = \quad \mathrm{ceil}\left(\frac{n_{\mathrm{samples}}}{r_{\mathrm{s}}} \cdot r_{\mathrm{sd}}\right). \tag{40}$$

6.2.7 Noise reduction

$\widehat{N}'_{\mathrm{tonal}}(l,z)$ is a first estimation of the specific loudness of the tonal component. However, the specific loudness of the tonal component is usually overestimated at this stage of the estimation process due to the tonal character of noise in the narrow-band filtered bands. Thus, further noise reduction is necessary. This is done by application of nonlinear sigmoid weighting of tonal vs. noise components. $\widehat{N}'_{\mathrm{tonal}}(l,z)$ is the tonal part of the specific loudness of the complete band-pass signal. The corresponding specific loudness of the complete band-pass signal is given by the autocorrelation function at zero lag:

$$N'_{\mathrm{signal}}(l,z) \quad = \bar{\varphi}'_{l,z}(m=0). \tag{41}$$

A first approximation of the signal-to-noise ratio in the band of interest can be derived as

$$\widehat{\mathrm{SNR}}(l,z) \quad = \quad \frac{\widehat{N}'_{\mathrm{tonal}}(l,z)}{N'_{\mathrm{signal}}(l,z) - \widehat{N}'_{\mathrm{tonal}}(l,z) + \varepsilon}. \tag{42}$$

Since the estimation of the tonal component might contain unsteady parts, low-pass filtering is performed over the temporal dimension of $\widehat{N}'_{\mathrm{tonal}}(l,z)$ and $\widehat{\mathrm{SNR}}(l,z)$. A cutoff frequency of 3,5 Hz is used.[20] Low-pass filters with the same filter coefficients are used for all critical bands. The filter defined in Formula (11) is used with order $k = 3$. The filter coefficients of the low-pass filter $h_{\mathrm{LP}}(l)$ can be calculated according to Formulae (14) and (15).[21] The filtered signals are then

$$\widetilde{N}'_{\mathrm{tonal}}(l,z) \quad = \quad \widehat{N}'_{\mathrm{tonal}}(l,z) * h_{\mathrm{LP}}(l) \tag{43}$$

and

$$\widetilde{\mathrm{SNR}}(l,z) \quad = \quad \widehat{\mathrm{SNR}}(l,z) * h_{\mathrm{LP}}(l), \tag{44}$$

where $*$ denotes the convolution. These filtered signals are used for further processing in Formulae (45) and (47).

Band-dependent noise reduction is achieved by weighting the filtered specific loudness $\widetilde{N}'_{\mathrm{tonal}}(l,z)$ of the tonal component by a sigmoid function

$$\mathrm{nr}(l,z) \quad = \quad \begin{cases} 1 - \mathrm{e}^{-\alpha\cdot\left(\frac{\widetilde{\mathrm{SNR}}(l,z)}{g(z)} - \beta\right)}, & \mathrm{e}^{-\alpha\cdot\left(\frac{\widetilde{\mathrm{SNR}}(l,z)}{g(z)} - \beta\right)} < 1 \\ 0 & \mathrm{e}^{-\alpha\cdot\left(\frac{\widetilde{\mathrm{SNR}}(l,z)}{g(z)} - \beta\right)} \geq 1 \end{cases}, \tag{45}$$

with parameters α and β as given in Table 7.

---

[20] Please note that the bandwidth of the lowpass filter is twice as large as the cut-off frequency! Therefore, the variable d in Formulae (14) and (15) should be calculated using $\tau(z) = \frac{1}{32} \cdot 6 \cdot \frac{1}{7\ \mathrm{Hz}} = 0.0268\mathrm{s}$ for all critical bands according to Formula (8).

[21] For Formula (15) the following factors e_i have to be used for a filter order of $k = 3$: $e_0 = 0, e_1 = 1, e_2 = 1$.

Table 7 — Parameters for the noise reduction function $\mathrm{nr}(l,z)$ (Formula (45))

Parameter	α	β
Value	20	0,07

Sigmoidal weighting significantly reduces wrongly-detected specific loudness of tonal components for broadband signals. The frequency dependent factor $g(z)$ is calculated as

$$g(z) = \frac{c(s_\mathrm{b}(z))}{F(z)^{d(s_\mathrm{b}(z))}}, \quad (46)$$

where the parameters c and d are given in Table 8 depending on the block size $s_\mathrm{b}(z)$ (see Table 4). This function mitigates frequency-dependent overestimations of the tonality estimation (due to the different block sizes) such that $\mathrm{SNR}(l,z)/g(z)$ is approximately constant over z for pink noise signals.

Table 8 — Parameters for the frequency dependent factor $g(z)$ (Formula (46))

$s_\mathrm{b}(z)$	8192	4096	2048	1024
$c(s_\mathrm{b}(z))$	18,21	12,14	417,54	962,68
$d(s_\mathrm{b}(z))$	0,36	0,36	0,71	0,69

The specific loudness of the tonal component, $N'_\mathrm{tonal}(l,z)$, is then modelled as

$$N'_\mathrm{tonal}(l,z) = \mathrm{nr}(l,z) \cdot \widetilde{N}'_\mathrm{tonal}(l,z). \quad (47)$$

6.2.8 Calculation of time-dependent specific tonality

The perceived tonality is not only dependent on the tonal content in each band, but also on the signal-to-noise ratio over all bands at each time instance l. Thus, to finally model the tonality of the signal, the overall loudness signal-to-noise ratio is evaluated across all bands. First, a new estimation of the specific loudness of the noise component is calculated, using the final estimation of the specific loudness of the tonal component:

$$N'_\mathrm{noise}(l,z) = N'_\mathrm{signal}(l,z) * h_\mathrm{LP}(l) - N'_\mathrm{tonal}(l,z). \quad (48)$$

The overall loudness signal-to-noise ratio is calculated as

$$\mathrm{SNR}(l) = \frac{\max_z N'_\mathrm{tonal}(l,z)}{\varepsilon + \sum_z N'_\mathrm{noise}(l,z)}. \quad (49)$$

A scaling factor

$$q(l) = \begin{cases} 1 - \mathrm{e}^{-A\cdot(\mathrm{SNR}(l)-B)}, & \mathrm{e}^{-A\cdot(\mathrm{SNR}(l)-B)} < 1 \\ 0 & \mathrm{e}^{-A\cdot(\mathrm{SNR}(l)-B)} \geq 1 \end{cases} \quad (50)$$

is applied multiplicatively. The parameters A and B are given in Table 9.

Table 9 — Parameters for the scaling factor (Formula (50))

Parameter	A	B
Value	35	0,003

Thus, the final estimation of the time-dependent specific tonality is given as:

$$T'(l,z) \quad = \quad c_{\mathrm{T}} \cdot q(l) \cdot N'_{\mathrm{tonal}}(l,z), \tag{51}$$

where $c_{\mathrm{T}} = 2{,}8785151$ is a calibration factor. The time index l can be mapped to the time t in seconds as:

$$t \quad = \quad \frac{l}{r_{\mathrm{sd}}} = \frac{l}{187{,}5}\ \mathrm{s}. \tag{52}$$

The unit of the tonality calculated by the psychoacoustic tonality method is given in $\mathrm{tu_{HMS}}$ (HMS stands for tonality units "according to the **H**earing **M**odel of **S**ottek" described in Clause 5). The psychoacoustic tonality method is calibrated using a 1 kHz tone with a sound pressure level of 40 dB. The tonality value shall be for this signal 1 $\mathrm{tu_{HMS}}$[22].

6.2.9 Calculation of averaged specific tonality

The specific tonality $T'(z)$ is taken by averaging the time-dependent specific tonality $T'(l,z)$. The averaging is performed as follows:

1. The first tonality values $T'(l,z)$ for $0 \leq l \leq 56$ (approximately corresponding to the first 300 ms of the input signal) are discarded due to the transient responses of the digital filters.
2. Only values that exceed a specific tonality value of 0,02 $\mathrm{tu_{HMS}}$ are used for averaging. This step ensures that the single value is independent of parts of the signal without noticeable tonal components.

This averaging can be described mathematically as

$$T'(z) \quad = \quad \frac{1}{\#(l'(z)) + \varepsilon} \sum_{l'} T'(l'(z), z)\,, \tag{53}$$

with

$$l'(z) \quad = \quad \{57 \leq l \leq l_{\mathrm{end}} \mid T'(l,z) > 0{,}02\ \mathrm{tu}_{HMS}\}, \tag{54}$$

using set notation[23]. The frequencies $f_{\mathrm{ton,z}}(z)$ are calculated by accordingly averaging the frequency $f_{\mathrm{ton}}(l,z)$ (see Formula (39)[24]) over corresponding time indices:

$$f_{\mathrm{ton,z}}(z) \quad = \quad \frac{1}{\#(l'(z)) + \varepsilon} \sum_{l'} f_{\mathrm{ton}}(l'(z), z). \tag{55}$$

6.2.10 Calculation of time-dependent tonality

The time-dependent tonality $T(l)$ is taken as the maximum of the time-dependent specific tonalities $T'(l,z)$ over all bands z. If the user is only interested in one specific tonal event, a user defined frequency range $[f_{\mathrm{L}}, f_{\mathrm{H}}]$ can be specified. In this case, only critical bands with the critical band number z are considered that fulfill the following requirements:

---

[22] The calibration factor c_{T} can be adjusted within a tolerance of 0,25 % to account for the effects of different implementations.

[23] In set notation, $\{x \mid \Phi(x)\}$ denotes all elements x with the property $\Phi(x)$. $\#(A)$ denotes the cardinality (i.e. the number of elements) of a set A.

[24] Note that $f_{\mathrm{ton}}(l,z)$ is denoted $f_{\mathrm{ton}}(z)$ in Eq. (39), since the time index l was neglected in this computation step.

$$16\ \text{Hz} \quad < \quad f_{\text{L}} \quad < \quad \frac{F(z) + F(z + 0{,}5)}{2} \tag{56}$$

and

$$20\ \text{kHz} \quad > \quad f_{\text{H}} \quad > \quad \frac{F(z) + F(z - 0{,}5)}{2} \tag{57}$$

leading to a range of critical bands between z_{L} and z_{H}. With this calculation procedure, the actually considered frequency range is $[f'_{\text{L}}, f'_{\text{H}}]$ with

$$f'_{\text{L}} \quad = \quad \min_{f} R(z_{\text{L}}). \tag{58}$$

and

$$f'_{\text{H}} \quad = \quad \min_{f} R(z_{\text{H}}). \tag{59}$$

with the frequency range $R(z)$

$$R(z) \quad = \quad \left[F(z) - \frac{\Delta f(z)}{2}, F(z) + \frac{\Delta f(z)}{2}\right]. \tag{60}$$

All frequency bands between z_{L} and z_{H}. are used for the maximum search:

$$T(l) \quad = \quad \max_{z \in [z_L, z_H]} T'(l, z). \tag{61}$$

The corresponding frequency $f_{\text{ton,l}}(l)$ is given as

$$f_{\text{ton,l}}(l) \quad = \quad f_{\text{ton}}(l, z_{\max}(l)). \tag{62}$$

where $z_{\max}(l)$ is the band in which the maximum of the time-dependent specific tonality $T'(l, z)$ was found for a given time instance l.

6.2.11 Calculation of representative values

The single value T of the tonality of the signal is taken by averaging the time-dependent overall tonality $T(l)$. The averaging is performed in the same way as described in Formula (53)

$$T \quad = \quad \frac{1}{\#(l')} \sum_{l'} T(l'), \tag{63}$$

with

$$l' \quad = \quad \{57 \leq l \leq l_{\text{end}} \mid T(l) > 0{,}02\ \text{tu}_{HMS}\}. \tag{64}$$

6.3 Information to be recorded for prominent tonalities

For stationary sounds, a tonal component in the critical band z_{tonal} is identified as prominent, if the specific tonality $T'(z_{\mathrm{tonal}})$ exceeds a value of 0,4 $\mathrm{tu_{HMS}}$ and the specific tonality has a local maximum in z_{tonal}. Additionally, the frequency $f_{\mathrm{ton},z}(z_{\mathrm{tonal}})$ needs to be in the range $[F(z_{\mathrm{tonal}}-1),\ F(z_{\mathrm{tonal}}+1)]$ for the component to be identified as prominent. If the user is only interested in one specific tonal event, a user defined frequency range $[f_{\mathrm{L}}, f_{\mathrm{H}}]$ can be specified. Then, only tonalities that are in the frequency range $[f'_{\mathrm{L}}, f'_{\mathrm{H}}]$[25] are considered: For each tonal component that has been identified as prominent according to this standard, the following information shall be recorded:

a) if a frequency range was defined, the resulting frequency range $[f'_{\mathrm{L}}, f'_{\mathrm{H}}]$ for searching prominent tonalities (Formulae (56) and (57));

b) the frequency, $f_{\mathrm{ton},z}(z_{\mathrm{tonal}})$, in hertz, of the tonality in the corresponding critical band z_{tonal} (see Formula (55));

c) details of the method used to evaluate the tonality (ECMA 418 – Part 2: Psychoacoustic metrics based on the hearing model – Clause 6.2 Psychoacoustic tonality calculation method), together with a reference to this Standard;

d) the psychoacoustic tonality value $T'(z_{\mathrm{tonal}})$ (see Formula (53)).

e) optionally, the time-dependent specific tonality $T'(l,z)$ (see Formula (51)).

For non-stationary sounds, a signal is considered to contain prominent tonalities, if the time-independent single value T of the time-dependent tonality $T(l)$[26] exceeds a value of 0,4 $\mathrm{tu_{HMS}}$ (see Formula (63)). If the signal has been identified to contain prominent tonalities according to this clause, the following information shall be recorded:

a) if a frequency range was defined, the resulting frequency range $[f'_{\mathrm{L}}, f'_{\mathrm{H}}]$ for searching prominent tonalities (Formulae (56) and (57));

b) the time-dependent frequency, $f_{\mathrm{ton,l}}(l)$, in hertz (see Formula (62)) of the time-dependent tonality $T(l)$;

c) details of the method used to evaluate the tonality (ECMA 418 – Part 2: Psychoacoustic metrics based on the hearing model – Clause 6.2 Psychoacoustic tonality calculation method), together with a reference to this Standard;

d) the time-dependent psychoacoustic tonality value $T(l)$ (see Formula (61));

e) the time-independent single value T (see Formula (63));

f) optionally: the time-dependent specific tonality $T'(l,z)$ (see Formula (51)).

NOTE The criterion for prominence of tonalities for the psychoacoustic tonality calculation method (Clause 6.2) is independent of frequency 0,4 $\mathrm{tu_{HMS}}$ (HMS stands for tonality units “according to the **H**earing **M**odel of **S**ottek” described in Clause 5).

[25] $[f'_{\mathrm{L}}, f'_{\mathrm{H}}]$ is calculated from $[f_{\mathrm{L}}, f_{\mathrm{H}}]$ as explained in Formulae (56) - (60).

[26] The time index l can be mapped to a time in seconds according to Formula (52).

7 Identification and evaluation of prominent roughness using a psychoacoustic roughness calculation method

This clause describes a perception-model-based procedure for determining whether or not noise emissions contain prominent roughness, and if present, their strengths: the psychoacoustic roughness calculation method. The calculation is based on the specific loudness as described in Clause 5.

The auditory sensation roughness describes, together with the auditory sensation fluctuation strength, the perception of temporal variations of sounds. While fluctuation strength covers slow variations (typically below 20 Hz), roughness is produced by faster variations up to around 500 Hz. The maximum of the auditory sensation is located at around 4 Hz modulation rate for fluctuation strength and 70 Hz modulation rate for roughness. Both auditory sensations can be produced either by amplitude modulation or by frequency modulation. Generally, periodic modulations produce higher values of fluctuation strength and roughness than stochastic variations.

Roughness is used for the subjective evaluation of sound characteristics as well as for sound design. With increasing roughness, sounds are increasingly attracting attention and perceived as increasingly aggressive, and annoying, without showing a difference in loudness or A-weighted sound pressure level.

The impression of roughness arises if a time-variant envelope is present in one critical band, for example tones with a temporal structure because of a change in amplitude or frequency. If these variations are rather slow (for example lower than 10 Hz), the auditory system is capable to follow the changes and a perception of fluctuation arises. With increasing modulation rates, sensations like R-roughness (around 20 Hz) arise and turn into actual roughness, where the auditory system is not capable of resolving the temporal variations. Variations of the envelope with modulation rates between 20 Hz and 300 Hz are perceived as “rough”. Roughness depends on the center frequency, the modulation rate f_{mod}, the degree of modulation m and the sound pressure level. Frequency modulated sounds produce a similar roughness as amplitude modulated sounds. The unit of roughness is “asper”. As reference signal with $R = 1$ asper, an amplitude modulated sinusoid of 1 kHz center frequency, $m = 1$, $f_{\mathrm{mod}} = 70$ Hz and a sound pressure level of 60 dB was chosen.

Roughness originates for example from a multiplicative combination of two vibrations – such as for example the gear mesh frequency and the rotational speed in a gear wheel – or from superposition of two or more tonal or narrowband sounds with a similar frequency. In practice, roughness often occurs in rotating components (engines, gearboxes, fans).

7.1 Psychoacoustic roughness calculation method

7.1.1 Overview

The psychoacoustic roughness calculation is based on scaled envelope power spectra $\Phi_{\mathrm{E},l,z}(k)$, which are calculated using the specific basis loudness $N'_{\mathrm{basis}}(l,z)$ (see Formula (25)) and the envelope of the $\mathrm{CBF} = 53$ segmented band-pass signals $p_{l,z}(n')$ (see Clause 5.1.5) as described in Clause 5. For the calculation of these values, a block size of $s_{\mathrm{b}} = 16384$ and a hop size of $s_{\mathrm{h}} = 4096$ for the segmentation in Clause 5.1.5 shall be used.

The further processing for roughness calculation is shown in Figure 8 and described in detail as follows:

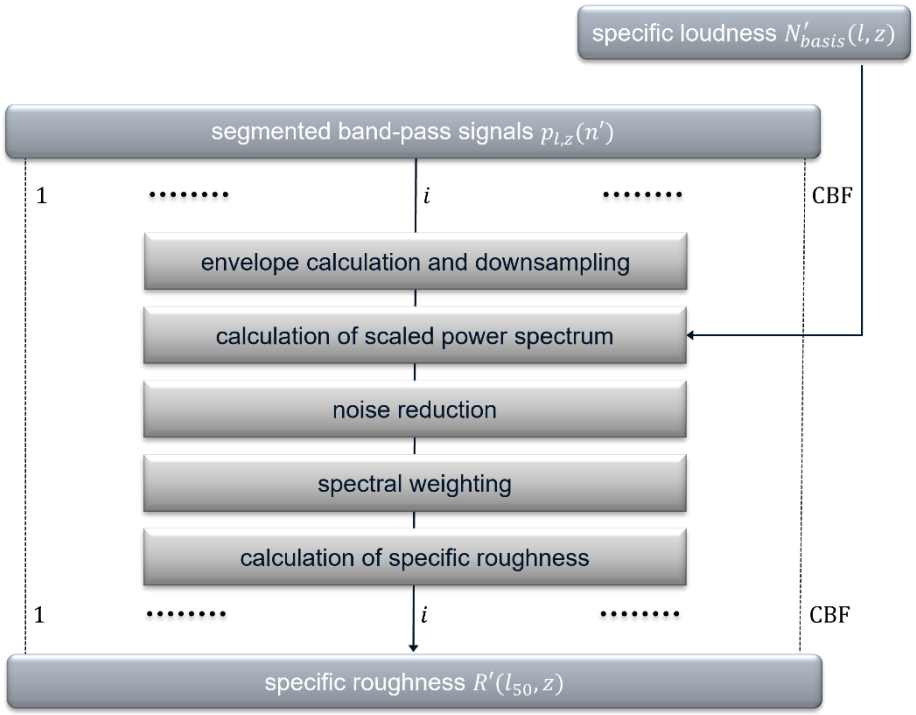

Figure 8 — Calculation of roughness based on band pass signals and the specific basis loudness calculated as described in Clause 5.

7.1.2 Envelope calculation and downsampling

The low-frequency envelopes are calculated from the segmented band-pass filtered sound pressure signals $p_{l,z}(n')$ (see Clause 5.1.5) using the Hilbert transform. The envelopes $p_{\mathrm{E},l,z}(n')$ are taken as magnitude of the analytical signals

$$p_{\mathrm{E},l,z}(n') \quad = \quad \left|p_{l,z}(n') + j\mathcal{H}(p_{l,z}(n'))\right|, \tag{65}$$

with $\mathcal{H}(\cdot)$ denoting the Hilbert transform. Since the envelopes only contain low modulation rates, downsampling with a factor of 32 is performed. The resulting downsampled envelopes of the band-pass signals are denoted $p_{\mathrm{E},l,z}(\tilde{n})$ [27]. With this step, the sampling rate changes from $r_\mathrm{s} = 48\,\mathrm{kHz}$ to $\tilde{r}_\mathrm{s} = 1500\,\mathrm{Hz}$. The block size $\tilde{s}_\mathrm{b} = 512$ and a hop size of $\tilde{s}_\mathrm{h} = 128$ are the values corresponding to the block size of $s_\mathrm{b} = 16384$ and the hop size of $s_\mathrm{h} = 4096$ for the segmentation in Clause 5.1.5.

[27] $\tilde{n}$ refers to the index of the downsampled signal.

7.1.3 Calculation of scaled power spectrum

The envelopes $p_{\mathrm{E},l,z}(\tilde{n})$ are windowed with a von-Hann window[28], $w_{\mathrm{Hann}}(\tilde{n})$) and a scaled power spectrum[29] $\Phi_{\mathrm{E},l,z}(k)$ is generated by using

$$\Phi_{\mathrm{E},l,z}(k) = \begin{cases} 0, & N'_{\max}(l)\cdot\varphi_{\mathrm{E},l,z}(0) = 0 \\ \left(N'_{\mathrm{basis}}(l,z)\right)^2 \cdot \left(\frac{\mathrm{Bark_{HMS}}}{\mathrm{sone_{HMS}}}\right) \Big| \mathrm{DFT}_{\tilde{s}_\mathrm{b}}\left(p_{\mathrm{E},l,z}(\tilde{n})\cdot w_{\mathrm{Hann}}(\tilde{n})\right)\Big|^2, & \text{else} \end{cases} \quad (66)$$

where $\mathrm{DFT}_{\tilde{s}_\mathrm{b}}$ denotes the $\tilde{s}_\mathrm{b}$-point Discrete Fourier Transform[30], k is the index corresponding to a modulation frequency of $k\cdot\frac{\tilde{r}_\mathrm{s}}{\tilde{s}_\mathrm{b}}$ Hz, $N'_{\mathrm{basis}}(l,z)$ is the specific basis loudness, $N'_{\max}(l) = \max_z\left(N'_{\mathrm{basis}}(l,z)\right)$ and

$$\varphi_{\mathrm{E},l,z}(0) = \sum_{\tilde{n}=0}^{\tilde{s}_\mathrm{b}-1}\left(p_{\mathrm{E},l,z}(\tilde{n})\cdot w_{\mathrm{Hann}}(\tilde{n})\right)^2. \quad (67)$$

The resulting quantities of $\Phi_{\mathrm{E},l,z}(k)$ are without units.

This step consideres the fact that the sensation of roughness changes nonlinearly with loudness. The results are scaled envelope power spectra $\Phi_{\mathrm{E},l,z}(k)$ which are used for further analysis of the roughness.

7.1.4 Noise reduction of the envelopes

Noise reduction of the envelopes is performed in two steps: First, the scaled power spectra of neighbouring bands are averaged to reduce noise effects. Averaging is performed over 3 bands. Each band is averaged with one higher and one lower band. This step results in averaged scaled power spectra $\bar{\Phi}_{\mathrm{E},l,z}(k)$.

Then, the sum of the averaged scaled power spectra,

$$s(l,k) = \sum_z \bar{\Phi}_{\mathrm{E},l,z}(k) \quad (68)$$

is calculated, showing an overview of all the modulation patterns over time. Each band may contain fluctuations even in the case of unmodulated noise due to the bandpass-filtering, but in this case the correlation between neighbouring bands is very low, while for modulated noise, the correlation is very high. The summation of the averaged scaled power spectra amplifies the correlated components (peaks) stronger than the uncorrelated ones. As a result, constant and/or time-varying peaks of the modulation spectrum become cleary visible. Now, the averaged scaled power spectra are weighted with a noise suppression weighting factor $w(l,k)$ depending on $s(l,k)$, that is applied to each individual critical band z, in order to distinguish between peaks related to the roughness perception and the background noise of the envelope.

$$\hat{\Phi}_{\mathrm{E},l,z}(k) = \bar{\Phi}_{\mathrm{E},l,z}(k)\cdot w(l,k) \quad (69)$$

---

[28] Here, the scaled von-Hann window is defined as $w_{\mathrm{Hann}}(\tilde{n}) = \frac{0{,}5-0{,}5\cos\left(\frac{2\pi\tilde{n}}{512}\right)}{\sqrt{0{,}375}}$. The scaling factor in the denominator ensures a correct estimation of the magnitude of the power spectrum.

[29] In the original version of the algorithm (22), the spectrum of the autocorrelation function $\varphi_{\mathrm{E},l,z}(m)$ of the envelope $p_{\mathrm{E},l,z}(n)$ was evaluated (m: lag time), corresponding to the power spectrum of the envelope. It should be noted that $\Phi_{\mathrm{E},l,z}(k)$ is not the Fourier transform of $\varphi_{\mathrm{E},l,z}(m)$ since the scaling of $\Phi_{\mathrm{E},l,z}(k)$ is not part of the autocorrelation function $\varphi_{\mathrm{E},l,z}(m)$.

[30] DFT of length N is defined as: $X(k) = \mathrm{DFT}_N\left(x(n)\right) = \sum_{n=0}^{N-1} x(n)\cdot \mathrm{e}^{-j2\pi kn/N}$ with $k = 0,1,\ldots,N-1$.

with the weigthing factor

$$w(l,k) \quad = \quad \begin{cases} \text{clip}(\widetilde{w}(l,k) - 0{,}1407, 0, 1)\,, & \widetilde{w}(l,k) \geq 0{,}05 \cdot \max\limits_{k=2,\ldots,255}(\widetilde{w}(l,k)) \\ 0, & \text{else} \end{cases} \tag{70}$$

where $\text{clip}(x, x_{\text{min}}, x_{\text{max}})$ returns clipped values of x between x_{min} and x_{max}. $\widetilde{w}(l,k)$ is calculated as

$$\widetilde{w}(l,k) \quad = \quad 0{,}0856 \cdot \frac{s(l,k)}{\tilde{s}(l) + \delta} \cdot \text{clip}(0{,}1891 \cdot \text{e}^{0{,}0120 \cdot k}, 0, 1) \tag{71}$$

with the median $\tilde{s}(l)$ of $s(l,k)$ over $k = 2, \ldots, 255$, and an additional exponential weighting depending on the modulation rate. The constant $\delta = 10^{-10}$ ensures a defined value of $\widetilde{w}(l,k)$ if $\tilde{s}(l) = 0$.

Note that for modulated signals the median value $\tilde{s}(l)$ is small compared to the peaks, whereas for unmodulated signals, $\tilde{s}(l)$ and the random peaks have almost the same magnitude, thus leading to large ratios $s(l,k)/\tilde{s}(l)$ for modulated signals; $w(l,k)$ tends to be 1, whereas for unmodulated signals $w(l,k)$ becomes 0. The parameters in the Formulae (70) and (71) were chosen that for an unmodulated White Gaussian Noise with a level of 80 dB all the weighting values $w(l,k)$ become 0, consequently leading to a roughness value of 0 asper.

7.1.5 Spectral weighting

In this step, the amplitudes of the averaged scaled power spectra are weighted according to the perception of roughness, which depends on the modulation rate. The spectral weighting is divided into four steps: First, spectral peaks are identified, and the modulation rate of those peaks is estimated with high precision. The amplitudes of peaks with a high modulation rate are weighted corresponding to the estimated modulation rate in the second step. Since usually, more than one peak is found, a third step is performed to analyse the relation of the different peaks. It is assumed that there is one dominant harmonic complex (a fundamental modulation rate with harmonics at multiples of the fundamental modulation rate) which is the dominant cause for roughness perception. The fundamental modulation rate of such a harmonic complex is estimated in the third step. In the fourth step, the amplitudes of peaks with a low modulation rate are weighted corresponding to the estimated fundamental modulation rate and summed to result in a first, uncalibrated estimation of the specific roughness.

7.1.5.1 Peak picking

In the peak picking steps, maxima of the averaged scaled power spectra are searched. To obtain a very precise estimation of the modulation rates corresponding to these maxima, a quadratic fit of the envelope spectrum is performed. Since the use of the von-Hann window in the calculation of the DFT does not lead to an exact quadratic shape in the spectrum, an additional refinement step is performed to reduce this bias.

First, local maxima of the averaged scaled power spectra $\widehat{\Phi}_{\text{E},l,z}(k)$ for $k = 2, \ldots, 255$ are searched. For each maximum, a corresponding prominence is calculated as the difference between the amplitude of the maximum and the surrounding values.To measure the prominence of a peak, a horizontal line is first extended from the peak to the left and right of the peak. The points where the line intersects the data on the left and right (this is either another peak or the end of the data) are marked as the outer endpoints of the left and right intervals. Next, the lowest valley is searched in both intervals. The larger of these two valleys is taken, and the vertical distance from that valley to the peak is measured. This distance is the prominence. Only the ten maxima with the highest prominence are considered. The maxima are numbered with i, where $i = 1$ is the maximum corresponding to the lowest modulation rate.

Only maxima at a modulation rate fulfilling the condition

$$\widehat{\Phi}_{\text{E},l,z}\left(k_{\text{p},i}(l,z)\right) > 0{,}05 \cdot \max_{i}\left(\widehat{\Phi}_{\text{E},l,z}\left(k_{\text{p},i}(l,z)\right)\right) \tag{72}$$

are considered, where $k_{\text{p},i}(l,z)$ desribes the modulation rate index k of the ith maximum.

Since the modulation rate index k only provides a limited resolution of the modulation rate, a refinement step is performed, which improves the spectral resolution of the estimated modulation rate and the corresponding

amplitudes of each peak. First, a quadratic fit coefficient vector $\mathbf{C} = (c_0, c_1, c_2)^T$ is calculated for each maximum, which contains three coefficients for a quadratic fit of the envelope spectrum around a centre modulation rate index. The vector is calculated by solving the system of equations

$$\hat{\mathbf{\Phi}}_{\mathrm{E},l,z} = \mathbf{K} \cdot \mathbf{C} \tag{73}$$

with

$$\hat{\mathbf{\Phi}}_{\mathrm{E},l,z} = \begin{pmatrix} \hat{\Phi}_{\mathrm{E},l,z}\left(k_{\mathrm{p},i}(l,z)-1\right) \\ \hat{\Phi}_{\mathrm{E},l,z}\left(k_{\mathrm{p},i}(l,z)\right) \\ \hat{\Phi}_{\mathrm{E},l,z}\left(k_{\mathrm{p},i}(l,z)+1\right) \end{pmatrix} \tag{74}$$

and the modulation index matrix for the quadratic fit,

$$\mathbf{K} = \begin{pmatrix} \left(k_{\mathrm{p},i}(l,z)-1\right)^2 & k_{\mathrm{p},i}(l,z)-1 & 1 \\ \left(k_{\mathrm{p},i}(l,z)\right)^2 & k_{\mathrm{p},i}(l,z) & 1 \\ \left(k_{\mathrm{p},i}(l,z)+1\right)^2 & k_{\mathrm{p},i}(l,z)+1 & 1 \end{pmatrix}. \tag{75}$$

From these coefficients, a first corrected modulation rate

$$\tilde{f}_{\mathrm{p},i}(l,z) = -\frac{c_1}{2c_0} \cdot \Delta f \tag{76}$$

is calculated with the DFT resolution $\Delta f = \frac{\tilde{r}_\mathrm{s}}{\tilde{s}_\mathrm{b}} = 1500\ \mathrm{Hz} / 512 = 2{,}9297\ \mathrm{Hz}$. The estimated modulation rate is refined by applying a bias correction term $\rho(\tilde{f}_{\mathrm{p},i}(l,z))$

$$f_{\mathrm{p},i}(l,z) = \tilde{f}_{\mathrm{p},i}(l,z) + \rho\left(\tilde{f}_{\mathrm{p},i}(l,z)\right). \tag{77}$$

The bias comes from approximating the spectrum of the von-Hann window with a quadratic function, when estimating the true modulation rate from the peaks in the sampled spectrum. The bias adjustment term depends almost only on the difference between the peak index and the corresponding exact modulation rate. This term $E(\theta)$ is calculated for 32 steps, covering a range of Δf, using integer steps $\theta = 0, \ldots, 32$ to indicate the corresponding sub-interval. A higher resolution of the modulation rate could be achieved by using more sub-intervals. Another option is the linear interpolation of $E(\theta)$ as a function of $\beta(\theta)$, the theoretical error after applying a correction, and θ_{corr}, the argument leading to the smallest error $\beta(\theta)$, as shown in the following:

$$\rho\left(\tilde{f}_{\mathrm{p},i}(l,z)\right) = E(\theta_{\mathrm{corr}}) - \left(E(\theta_{\mathrm{corr}}) - E(\theta_{\mathrm{corr}}-1)\right) \cdot \frac{\beta(\theta_{\mathrm{corr}}-1)}{\beta(\theta_{\mathrm{corr}}) - \beta(\theta_{\mathrm{corr}}-1)} \tag{78}$$

θ_{corr} is determined from the set of possible integer θ values that lie between 0 and 32 (the value of $\theta = 33$ in Table 10 is given only to simplify the implementation, to avoid the use of additional conditions in Formula (81)). For each possible value of θ, $\beta(\theta)$ is calculated from:

$$\beta(\theta) = \left(\mathrm{floor}\left(\frac{\tilde{f}_{\mathrm{p},i}(l,z)}{\Delta f}\right) + \frac{\theta}{32}\right) \cdot \Delta f - \left(\tilde{f}_{\mathrm{p},i}(l,z) + E(\theta)\right) \tag{79}$$

where $\mathrm{floor}(x)$ gives the greatest integer value smaller than or equal to the number x. θ_{min} is the θ value that produces the smallest beta value magnitude:

$$\theta_{\mathrm{min}} = \underset{0 \le \theta \le 32}{\mathrm{argmin}} |\beta(\theta)|. \tag{80}$$

θ_{corr} is then calculated from:

$$\theta_{\mathrm{corr}} = \begin{cases} \theta_{\min}, & \theta_{\min} > 0 \text{ and } \beta(\theta_{\min}) \cdot \beta(\theta_{\min} - 1) < 0 \\ \theta_{\min} + 1, & \text{else} \end{cases}. \tag{81}$$

Table 10 and Formula (79) are used to calculate the parameters needed to calculate the bias term given in Formula (78).

Table 10 – Error correction values $E(\theta)$

θ	0	1	2	3	4	5	6	7	8
$E(\theta)$/Hz	0,0000	0,0457	0,0907	0,1346	0,1765	0,2157	0,2515	0,2828	0,3084
θ	9	10	11	12	13	14	15	16	17
$E(\theta)$/Hz	0,3269	0,3364	0,3348	0,3188	0,2844	0,2259	0,1351	0,0000	-0,1351
θ	18	19	20	21	22	23	24	25	26
$E(\theta)$/Hz	-0,2259	-0,2844	-0,3188	-0,3348	-0,3364	-0,3269	-0,3084	-0,2828	-0,2515
θ	27	28	29	30	31	32	33		
$E(\theta)$/Hz	-0,2157	-0,1765	-0,1346	-0,0907	-0,0457	0,0000	0,0000		

The amplitudes of the maxima are calculated as

$$A_i(l,z) = \sum \widehat{\mathbf{\Phi}}_{\mathrm{E},l,z} = \sum_{m=-1}^{1} \widehat{\Phi}_{\mathrm{E},l,z}\left(k_{\mathrm{p},i}(l,z) + m\right), \tag{82}$$

where it is assumed that the energy of a peak is mainly distributed over the index of the maximum and the two neighbouring indices due to the use of the von-Hann window in the DFT calculation.

7.1.5.2 Weighting of high modulation rates

In a next step, these amplitudes are weighted with a modulation-rate-dependent factor $G_{l,z,i}(f_{\mathrm{p},i}(l,z))$ and a scaling factor $r_{\max}(z)$. This weighting (together with the weigthing of low modulation rates described in 7.1.5.4) consideres the dependency of the perceived roughness on the modulation rate. The weighting parameters were obtained by an optimization procedure, fitting the results of the roughness algorithm to the results of listening tests for sinusoids of different carrier frequencies with different modulation rates from Reference [12]. Those results are shown in the evaluation of the roughness algorithm in Annex C, Figure C.1 and also in Reference [30].

$$\tilde{A}_i(l,z) = \begin{cases} A_i(l,z) \cdot r_{\max}(z), & f_{\mathrm{p},i}(l,z) < f_{\max}(z) \\ G_{l,z,i}(f_{\mathrm{p},i}(l,z)) \cdot A_i(l,z) \cdot r_{\max}(z), & f_{\mathrm{p},i}(l,z) \geq f_{\max}(z) \end{cases} \tag{83}$$

with

$$r_{\max}(z) = \frac{1}{1 + r_1 \left| \log_2\left(\frac{F(z)}{1\,\mathrm{kHz}}\right) \right|^{r_2}} \tag{84}$$

and the corresponding parameters r_1 and r_2 as given in Table 11.

Table 11 – Parameters for $r_{\max}(z)$

	$F(z) < 1$ kHz	$F(z) \geq 1$ kHz
r_1	0,3560	0,8024
r_2	0,8049	0,9333

The weighting factor $G_{l,z,i}(f_{\mathrm{p},i}(l,z))$ is calculated as

$$G_{l,z,i}(f_{\mathrm{p},i}(l,z)) = \frac{1}{\left(1+\left(\left(\frac{f_{\mathrm{p},i}(l,z)}{f_{\max}(z)}-\frac{f_{\max}(z)}{f_{\mathrm{p},i}(l,z)}\right)\cdot q_1\right)^2\right)^{q_2(z)}} \tag{85}$$

where

$$f_{\max}(z) = 72{,}6937\cdot\left(1-1{,}1739\cdot e^{-5{,}4583\cdot\frac{F(z)}{1\,\mathrm{kHz}}}\right)\ \mathrm{Hz} \tag{86}$$

is the modulation rate at which the weighting factor reaches the maximum of one. $F(z)$ is the center frequency of the auditory filter bank as descibed in Clause 5. The parameter $q_1 = 1{,}2822$ and $q_2(z)$ is calculated as

$$q_2(z) = \begin{cases} 0{,}2471, & \frac{F(z)}{1\,\mathrm{kHz}} < 2^{-3{,}4253} \\ 0{,}2471+0{,}0129\cdot\left(\log_2\left(\frac{F(z)}{1\,\mathrm{kHz}}\right)+3{,}4253\right)^2, & \frac{F(z)}{1\,\mathrm{kHz}} \geq 2^{-3{,}4253} \end{cases} \tag{87}$$

7.1.5.3 Estimation of fundamental modulation rate

In this step, the maxima of the averaged scaled power spectra, which were found in 7.1.5.1 are further analysed. It is assumed that there is one dominant harmonic complex (a fundamental modulation rate with harmonics at multiples of the fundamental modulation rate) which is the dominant cause for roughness perception. The fundamental modulation rate of such a harmonic complex is estimated in this step.

For each block l and band z, the fundamental modulation rate of the envelope is estimated in the next processing step considering the modulation rate $f_{\mathrm{p},i}(l,z)$ and the amplitude $\tilde{A}_i(l,z)$ of the block. Since the dependencies on l and z are not relevant for this processing step, the variables will be denoted only in dependency of the index of the corresponding maximum, i, $f_{\mathrm{p}}(i)$ and $\tilde{A}(i)$ in the following to simplify the notation.

For each maximum with index i, it is tested whether the corresponding modulation rate $f_{\mathrm{p}}(i)$ is the best estimate for the fundamental modulation rate of the envelope, by assuming that the sum over the harmonic complex corresponding to the best estimate will result in the highest value. The excact procedure is described in the following, where i_0 describes the index of the currently tested maximum.

First, integer ratios of the modulation rates $f_{\mathrm{p}}(i)$ of all found maxima to the modulation rate $f_{\mathrm{p}}(i_0)$ are calculated

$$R_{i_0}(i) = \mathrm{round}\left(\frac{f_{\mathrm{p}}(i)}{f_{\mathrm{p}}(i_0)}\right), \tag{88}$$

by rounding to the nearest integer. If several i result in the same integer ratio $R_{i_0}(i)$, it needs to be decided which of the maxima is used further. In this case, the maximum with the index

$$i = \underset{i}{\mathrm{argmin}}\left|\frac{f_{\mathrm{p}}(i)}{R_{i_0}(i)\cdot f_{\mathrm{p}}(i_0)}-1\right| \tag{89}$$

is used, while the other maxima are discarded. From all remaining maxima, a set I_{i_0} of indices of all maxima, which belong to a harmonic complex with fundamental modulation rate $f_{\mathrm{p}}(i_0)$ is defined (using a tolerance of 4%):

$$I_{i_0} = \left\{i\ \middle|\ \left(\left|\frac{f_{\mathrm{p}}(i)}{R_{i_0}(i)\cdot f_{\mathrm{p}}(i_0)}-1\right|<0{,}04\right)\right\}. \tag{90}$$

For this set of indices, the energy of the harmonic complex is calculated as

$$E_{i_0} = \sum_{i \in I_{i_0}} \tilde{A}(i). \tag{91}$$

The index i_0 leading to the highest energy is denoted $i_{\max}$ in the following, the corresponding set of indices I_{i_0} is denoted $I_{\max}$. The fundamental modulation rate of the envelope is $f_{\mathrm{p}}(i_{\max})$.

In the following, only peaks corresponding to the indices in $I_{\max}$ are considered as part of the envelope. The amplitudes of these peaks are weighted depending on the distance between the center of gravity of these peaks and the modulation rate of the peak with the highest amplitude:

$$\hat{A}(i) = w_{\mathrm{peak}} \cdot \tilde{A}(i) \tag{92}$$

with $i \in I_{\max}$ and

$$w_{\mathrm{peak}} = 1 + 0{,}1 \cdot \left| \frac{\sum_{i \in I_{\max}} \left(\frac{f_{\mathrm{p}}(i)}{\mathrm{Hz}} \cdot \tilde{A}(i) \right)}{\sum_{i \in I_{\max}} \tilde{A}(i)} - \frac{f_{\mathrm{p}}(i_{\mathrm{peak}})}{\mathrm{Hz}} \right|^{0{,}749} \tag{93}$$

and

$$i_{\mathrm{peak}} = \underset{i \in I_{\max}}{\operatorname{argmax}} \tilde{A}(i). \tag{94}$$

7.1.5.4 Weighting of low modulation rates

In this next step, another weighting based on the fundamental modulation rate and a summation of amplitudes is performed. The block index l and the band index z are reintroduced for this step. Thus, the weighted amplitudes are denoted $\hat{A}_i(l,z)$, the corresponding fundamental modulation rates $f_{\mathrm{p},i_{\max}}(l,z)$ and the set of relevant maxima $I_{\max}(l,z)$.

The summation and weighting is performed as

$$A(l,z) = \begin{cases} \sum_{i \in I_{\max}(l,z)} G_{l,z,i}(f_{\mathrm{p},i_{\max}}(l,z)) \cdot \hat{A}_i(l,z), & f_{\mathrm{p},i_{\max}}(l,z) < f_{\max}(z) \\ \sum_{i \in I_{\max}(l,z)} \hat{A}_i(l,z), & f_{\mathrm{p},i_{\max}}(l,z) \geq f_{\max}(z) \end{cases} \tag{95}$$

where $G_{l,z,i}(f_{\mathrm{p},i_{\max}}(l,z))$ is calculated as described in Formula (85) but with parameters $q_1 = 0{,}7066$ and

$$q_2(z) = 1{,}0967 - 0{,}0640 \cdot \log_2\left(\frac{F(z)}{1\ \mathrm{kHz}}\right). \tag{96}$$

The parameter $f_{\max}(z)$ in Formula (95) is calculated according to Formula (86).

Values of $A(l,z)$ that fall below a threshold of $0{,}074376$ are set to zero.

7.1.6 Optional entropy weighting based on randomness of modulation rate

In an optional processing step, $A(l,z)$ is weighted depending on the randomness (measured using the entropy) of the estimated modulation rates. This method has been shown to improve the estimation of the roughness[31][32].

For this processing step, a signal of rotational speed $d(n)$ (unit revolutions per minute) as reference variable with the same sampling rate as the sound pressure signal $p(n)$ needs to be available.

First, the rotational speed signal is segmented in the same way as the sound pressure signal (see Clause 5.1.5, with s_b and s_h as given in Clause 7.1.1). The result is a segmented rotational speed signal $d_\mathrm{S}(n', l)$. In each time block l, the median of $d_\mathrm{S}(n', l)$ over n' is calculated. The result $\tilde{d}_\mathrm{S}(l)$ is an estimation of one rotational speed value for each block. This estimation is transformed to an estimation of the frequency of the rotational speed in Hertz:

$$f_\mathrm{D}(l) = \frac{\tilde{d}_\mathrm{S}(l)}{60\ \frac{\mathrm{R}}{\mathrm{min}}}\ \mathrm{Hz} \tag{97}$$

Now the maxima of the modulation rate, which were found in Clause 7.1.5.1 to calculate a weighting factor based on the entropy of these maxima. First, a set

$$I_\mathrm{f}(l, z) = \{i \mid i \notin I_\mathrm{max}(l, z) \vee i = i_\mathrm{max}\} \tag{98}$$

Is defined. This set contains all indices of maxima, which were not identified as corresponding to the harmonic complex of the estimated fundamental frequency in Section 7.1.5.3, and the index corresponding to the fundamental frequency (but not the ones of the harmonics). For all $i \in I_\mathrm{f}(l, z)$ an estimation of the order is calculated as the ratio between the frequency of the maximum $f_{\mathrm{p},i}(l, z)$ (see Clause 7.1.5.1) and the frequency of the rotational speed:

$$o_i(l, z) = \begin{cases} 0, & f_\mathrm{D}(l) = 0 \\ \dfrac{f_{\mathrm{p},i}(l, z)}{f_\mathrm{D}(l)}, & \text{else} \end{cases}. \tag{99}$$

Now a histogram of all estimated orders is calculated for each time index l and frequency band z from all maxima of the current time block and the three preceding and subsequent blocks[31]. In these histograms, 160 classes of constant width are used between the values 0,0625 and 20,625. The result is the histogram $H(b, l, z)$, where b is the class number and $H(b, l, z)$ contains the number of elements in the respective class. For calculation of the entropy, probabilities of occurrence

$$P(b, l, z) = \begin{cases} 0, & \sum_b H(b, l, z) = 0 \\ \dfrac{H(b, l, z)}{\sum_b H(b, l, z)}, & \text{else} \end{cases} \tag{100}$$

are calculated from the histogram for all classes. From this probability, the Shannon entropy

$$E(l, z) = \begin{cases} 0, & P(b, l, z) = 0 \\ -\sum_b \left(P(b, l, z) \cdot \log_2 P(b, l, z)\right), & \text{else} \end{cases}, \tag{101}$$

is calculated[32]. Finally, $A(l, z)$ is weighted with the entropy, if $E(l, z) > 1$:

$$A_\mathrm{E}(l, z) = \frac{A(l, z)}{\max(E(l, z); 1)}. \tag{102}$$

---

31 In the border regions less preceding or subsequent blocks are used.

32 In the case of a probability of zero, a result of $0 \cdot \log_2 0 = 0$ is used according to the limit $\lim_{x \to 0}(x \log_2 x) = 0$.

If this optional weighting step is used, $A_{\mathrm{E}}(l,z)$ needs to be used instead of $A(l,z)$ in all following processing steps of this algorithm.

7.1.7 Calculation of time-dependent specific roughness

$A(l,z)$ is interpolated to a sampling rate of $r_{\mathrm{s50}} = 50$ Hz using a piecewise cubic Hermitian function (temporal resolution of 20 ms). The new time index is designated l_{50}. Subsequently, negative values resulting from the interpolation are set to zero, resulting in a first, uncalibrated estimate of the specific roughness $R'_{\mathrm{est}}(l_{50},z)$.

Here, the results belonging to the zero-padding done at the start of the processing need to be removed. Thus the last evaluated block shall be:

$$l_{50,\mathrm{end}} \quad = \quad \mathrm{ceil}\left(\frac{n_{\mathrm{samples}}}{r_{\mathrm{s}}} \cdot r_{\mathrm{s50}}\right). \tag{103}$$

The next step in calculating the specific roughness is a nonlinear transform, depending on the distribution of $R'_{\mathrm{est}}(l_{50},z)$ over the critical bands z. This step is necessary to take into account that the roughness perception differs for broad-band signals (i.e., signals with a broader distribution of $R'_{\mathrm{est}}(l_{50},z)$ over the critical bands) compared to narrow band signals such as modulated sinusoids (i.e., signals with a narrow distribution of $R'_{\mathrm{est}}(l_{50},z)$ over the critical bands). With this step it is possible to model the roughness for very different kinds of synthetical and technical sounds as described in Reference [30].

Together with the nonlinear transform, a calibration is performed, which ensures that the calibration signal (amplitude modulated sinusoid, 60 dB SPL, 1 kHz carrier frequency, 70 Hz modulation rate) results in a roughness of 1 asper[33].

$$\hat{R}'(l_{50},z) \quad = \quad c_{\mathrm{R}} \cdot \left(R'_{\mathrm{est}}(l_{50},z)\right)^{E(l_{50})} \tag{104}$$

with the calibration factor $c_{\mathrm{R}} = 0{,}0180909 \ \frac{\mathrm{asper}}{\mathrm{Bark_{HMS}}}$,

$$E(l_{50}) \quad = 0{,}95555 \cdot \left(\tanh\left(1{,}6407 \cdot \left(B(l_{50}) - 2{,}5804\right)\right) + 1\right) \cdot 0{,}5 + 0{,}58449 \tag{105}$$

and

$$B(l_{50}) \quad = \begin{cases} \dfrac{\tilde{R}'_{\mathrm{est}}(l_{50})}{\bar{R}'_{\mathrm{est}}(l_{50})}, & \bar{R}'_{\mathrm{est}}(l_{50}) \neq 0 \\ 0, & \bar{R}'_{\mathrm{est}}(l_{50}) = 0 \end{cases} \tag{106}$$

The squared and linear mean $\tilde{R}'_{\mathrm{est}}(l_{50})$ and $\bar{R}'_{\mathrm{est}}(l_{50})$ are defined as

$$\tilde{R}'_{\mathrm{est}}(l_{50}) \quad = \quad \sqrt{\frac{\sum_z \left(R'_{\mathrm{est}}(l_{50},z)\right)^2}{\mathrm{CBF}}}, \tag{107}$$

and

$$\bar{R}'_{\mathrm{est}}(l_{50}) \quad = \quad \frac{\sum_z \left(R'_{\mathrm{est}}(l_{50},z)\right)}{\mathrm{CBF}} \tag{108}$$

where $\mathrm{CBF} = 53$ is the number of critical bands. The resulting estimate of the time-dependent specific roughness, $\hat{R}'(l_{50},z)$, is smoothed by using a lowpass filter of order one with different time constants for rising and falling slopes. This filtering consideres the fact, that the perception of sound events rises quickly with the

[33] The calibration factor c_{R} can be adjusted within a tolerance of 0,25 % to account for the effects of different implementations.

beginning of the sound event, but only decays slowly when the sound event ends. A similar filtering is used in the loudness model for time-varying sounds of Moore and Glasberg [6]. The filtering can be described as

$$R'(l_{50},z) = \begin{cases} \hat{R}'(l_{50},z), & l_{50}=0 \\ \hat{R}'(l_{50},z)\cdot\left(1-\mathrm{e}^{-\frac{1}{r_{\mathrm{ss0}}\cdot\tau(l_{50},z)}}\right)+\hat{R}'(l_{50}-1,z)\cdot\mathrm{e}^{-\frac{1}{r_{\mathrm{ss0}}\cdot\tau(l_{50},z)}}, & l_{50}\geq 1 \end{cases} \quad (109)$$

with the different time constants for rising and falling slopes

$$\tau(l_{50},z) = \begin{cases} 0{,}0625, & \hat{R}'(l_{50},z)\geq\hat{R}'(l_{50}-1,z) \\ 0{,}5000, & \hat{R}'(l_{50},z)<\hat{R}'(l_{50}-1,z) \end{cases}, \quad (110)$$

resulting in the final estimate of the time-dependent specific roughness $R'(l_{50},z)$.

7.1.8 Calculation of representative values

The specific roughness $R'(z)$ is taken by averaging the time-dependent specific roughness $R'(l_{50},z)$. For the averaging, the first roughness values $R'(l_{50},z)$ for $0\leq l_{50}\leq 15$ (approximately corresponding to the first 300 ms of the input signal) are discarded due to the transient responses of the digital filters.

7.1.9 Calculation of time-dependent roughness

The time-dependent roughness $R(l_{50})$ is the integral of $R'(l_{50},z)$ over z, approximated by summing over all bands z while considering the overlap Δz:

$$R(l_{50}) = \Delta z\sum_{z}(R'(l_{50},z)). \quad (111)$$

7.1.10 Calculation of representative values

The single value R is calculated by taking the 90th percentile of the time-dependent roughness $R(l_{50})$, discarding again the first roughness values $R(l_{50})$ for $0\leq l_{50}\leq 15$.

7.1.11 Calculation of roughness for binaural signals

For binaural signals, monaural time-dependent specific roughness values $R'_{\mathrm{L}}(l_{50},z)$ and $R'_{\mathrm{R}}(l_{50},z)$ of the left and right channel shall be calculated separately for each channel (assuming diotic signals).

A combined binaural time-dependent specific roughness $R'_{\mathrm{B}}(l_{50},z)$ is calculated using the quadratic mean:

$$R'_{\mathrm{B}}(l_{50},z) = \sqrt{\frac{\left(R'_{\mathrm{L}}(l_{50},z)\right)^2+\left(R'_{\mathrm{R}}(l_{50},z)\right)^2}{2}}. \quad (112)$$

Formula (112) approximately corresponds to the formula for binaural inhibition from the binaural loudness model by Moore/Glasberg (ISO 532-2[7], see also Reference [33]). In the case that the roughness value of a channel is negligible, Formula (112) results in a roughness, which is $\sqrt{0{,}5}$ lower than that of the diotic presentation.

For binaural signals, the binaural time-dependent specific roughness $R'_{\mathrm{B}}(l_{50},z)$ shall be used as basis for the calculation of the specific roughness $R'(z)$, the time-dependent roughness $R(l_{50})$ and the single value R instead of $R'(l_{50},z)$ in Clauses 7.1.8, 7.1.9 and 7.1.10.

7.2 Information to be recorded for prominent roughness

A signal is considered to contain prominent roughness, if the time-independent single value R of the time-dependent roughness $R(l_{50})$ exceeds a value of 0,2 asper. If the signal has been identified to contain prominent roughness according to this standard, the following information shall be recorded:

a) details of the method used to evaluate the roughness (ECMA 418 – Part 2: Psychoacoustic metrics based on the hearing model – Clause 7.1 Psychoacoustic roughness calculation method), together with a reference to this Standard;

b) the time-dependent psychoacoustic roughness values $R(l_{50})$ (see Formula (111));

c) the time-independent single value R;

d) information if the optional entropy weighting was used or not;

e) optionally: the time-dependent specific roughness $R'(l_{50}, z)$.

8 Improved identification and evaluation of loudness using psychoacoustic methods of tonal and noise loudness

This clause describes a procedure based on a perceptual model for determining how loud a sound is perceived taking into consideration how people's perceptions differ for tonal and noise signals. For narrowband signals with subcritical bandwidths, it is generally assumed that loudness only depends on the level, independent of the bandwidth. This assumption is also demonstrated by standardized loudness models such as ISO 532-1 (Zwicker) [3] and ISO 532-3 DIS (Moore, Glasberg, Schlittenlacher). Several published experimental studies [35]-[38], however, showed that this is not the case, but rather that tonal components are perceived as louder than equivalent narrow-band (subcritical bandwidth) noise with the same level on the same band. Sottek et al [39] have shown that a more accurate loudness estimation can be done by combining the tonal loudness and noise loudness presented earlier. This calculation process is described in this Clause.

8.1 Psychoacoustic loudness calculation method

The calculation process is simpler compared to the last sections, since most of the calculations were already described in Clauses 5 and 6. An overview of the determination of the specific loudness is shown in Figure 8.

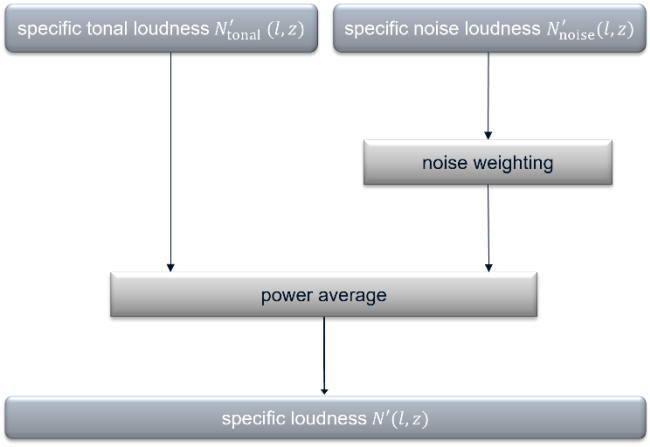

Figure 8 — Calculation of loudness based on specific tonal and noise loudness (see Clause 6).

8.1.1 Calculation of time-dependent specific loudness

To obtain a better estimation of the loudness, a power average of the specific tonal loudness $N'_{\mathrm{tonal}}(l,z)$ in Formula (47) and the weighted specific noise loudness $N'_{\mathrm{noise}}(l,z)$ in Formula (48) is performed to obtain the specific loudness $N'(l,z)$:

$$N'(l,z) = \left(\left(N'_{\mathrm{tonal}}(l,z) \right)^{e(z)} + \left(w_{\mathrm{n}} \cdot N'_{\mathrm{noise}}(l,z) \right)^{e(z)} \right)^{1/e(z)} \tag{113}$$

Here $w_{\mathrm{n}} = 0{,}5331$ and the exponent $e(z)$ is a function of the maximal specific basis loudness:

$$e(z) = \frac{a}{\max\limits_{z} \left(\frac{N'_{\mathrm{tonal}}(l,z) + N'_{\mathrm{noise}}(l,z)}{\left(\frac{\mathrm{sone}_{\mathrm{HMS}}}{\mathrm{Bark}_{\mathrm{HMS}}} \right)} \right) + \epsilon} + b \tag{114}$$

with the parameters a, b and ϵ given in Table 12.

Table 12 – Parameters to define the exponent for the loudness power average (Formula (114))

Parameter	a	b	ϵ
Value	0,2918	0,5459	10^{-12}

8.1.2 Calculation of averaged specific loudness

The specific loudness $N'(z)$ is taken by averaging the time-dependent specific loudness $N'(l,z)$. For the averaging, the first loudness values $N'(l,z)$ for $0 \leq l \leq 56$ (approximately corresponding to the first 300 ms of the input signal) are discarded due to the transient responses of the digital filters.

This averaging can be described mathematically as:

$$N'(z) = \left(\frac{1}{l_{\text{end}} - 56}\sum_{l} N'(l,z)^e\right)^{1/e}, \quad (115)$$

where $e = \frac{1}{\log_{10}(2)}$ and $57 \leq l \leq l_{\text{end}}$. A power average is used here because it gives more weight to stronger components and correlates better with human loudness perception[39]

8.1.3 Calculation of time-dependent loudness

The time-dependent loudness $N(l)$ is calculated by integrating all specific loudness values, like Formula (26) with $\Delta z = 0{,}5$:

$$N(l) = \sum_{i=1}^{\text{CBF}} N'\left(l, \frac{i}{2}\right) \cdot \Delta z. \quad (116)$$

The unit of the result is $\text{sone}_{\text{HMS}}/\text{Bark}_{\text{HMS}}$ and no additional calibration is needed since the specific results were already calibrated in Formula (23).

8.1.4 Calculation of representative values

The single value N of the loudness of the signal is taken again by a power average of the time-dependent loudness $N(l)$. Like the specific loudness, the values of $N(l)$ for $0 \leq l \leq 56$ (approximately corresponding to the first 300 ms of the input signal) are discarded due to the transient responses of the digital filters.

$$N = \left(\frac{1}{l_{\text{end}} - 56}\sum_{l} N(l)^e\right)^{1/e}, \quad (117)$$

where $l > 56$. The unit of the result is sone_{HMS}. While this process does not significantly modify the loudness of pure tonal signals in comparison to the result in Formula (26), it improves the result of noise-like signals and mixtures of tones and noise for which the loudness of the noise components are overestimated [39].

8.1.5 Calculation of loudness for binaural signals

For binaural signals, monaural time-dependent specific loudness values $N'_{\mathrm{L}}(l,z)$ and $N'_{\mathrm{R}}(l,z)$ of the left and right channel shall be calculated separately for each channel (assuming diotic signals).

A combined binaural time-dependent specific loudness $N'_{\mathrm{B}}(l,z)$ is calculated using the quadratic mean:

$$N'_{\mathrm{B}}(l,z) \quad = \quad \sqrt{\frac{\left(N'_{\mathrm{L}}(l,z)\right)^2 + \left(N'_{\mathrm{R}}(l,z)\right)^2}{2}}. \tag{118}$$

Formula (118) approximately corresponds to the formula for binaural inhibition from the binaural loudness model by Moore/Glasberg (ISO 532-2 [7], see also Reference [33]). In the case that the loudness value of a channel is negligible, Formula (118) results in a loudness, which is $\sqrt{0{,}5}$ lower than that of the diotic presentation.

For binaural signals, the binaural time-dependent specific loudness $N'_{\mathrm{B}}(l,z)$ shall be used as basis for the calculation of the specific loudness $N'(z)$, the time-dependent loudness $N(l)$ and the single value N instead of $N'(l,z)$ in Clauses 8.1.2, 8.1.3 and 8.1.4.

8.2 Information to be recorded for loudness

The following information shall be recorded:

a) details of the method used to evaluate the loudness (ECMA 418 – Part 2: Psychoacoustic metrics based on the hearing model – Clause 8.1 Psychoacoustic loudness calculation method), together with a reference to this Standard;

b) the time-dependent psychoacoustic loudness values $N(l)$ (see Formula (116));

c) the time-independent single value N;

d) optionally: the time-dependent specific loudness $N'(l,z)$.

Annex A
(informative)

Evaluation of the psychoacoustic hearing model

The psychoacoustic loudness calculation is evaluated by comparison with the target equal-loudness contours as shown in Figure 2. The loudness was calculated for sinusoidal signals with a frequency of 1000 Hz and a sound pressure level of 20 to 80 dB with a step size of 20 dB. For other frequencies, the level was varied to match the loudness calculated for the 1000 Hz tone. The same procedure was performed for the lower threshold of hearing. The results are shown in Figure A.1. The target equal-loudness contours are emulated well by the results of the hearing model.

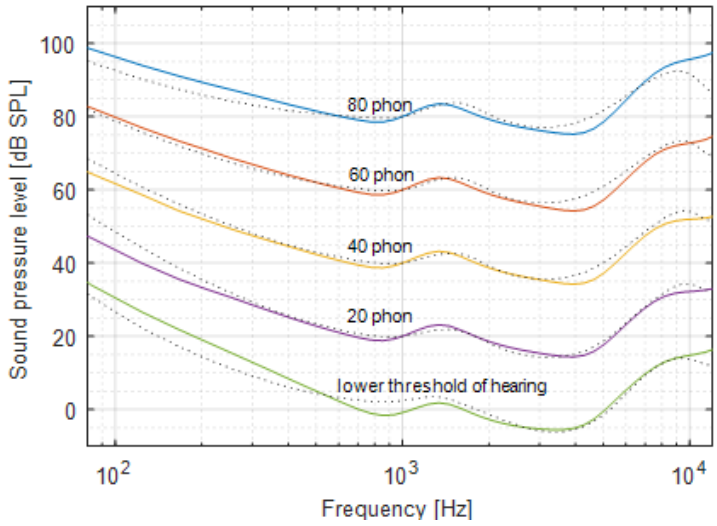

Figure A.1— Results for the equal-loudness contours. The dotted lines show the target equal-loudness contours, the solid lines are the equal-loudness contours obtained with the hearing model

Annex B
(informative)

Evaluation of the psychoacoustic tonality calculation method

B.1 Application examples

Figure B.1 shows analysis results for a frequency-modulated signal with a low modulation rate of 2 Hz, a modulation index of 150 at a frequency of 2 kHz and with very low sound pressure (L = 30 dB).

From top to bottom, it shows:

1. the spectrum (FFT size 65536, sampling rate 48 kHz), a smoothed spectrum (1/24th octave smoothed FFT: the “background noise”, useful to show general shapes while not resolving pure tones), and a 1-critical-bandwidth peak-hold spectrum as “critical bandwidth ruler”;
2. the tone-to-noise ratio [34] (TNR) results along with the TNR tolerance line.
3. the prominence ratio [35] (PR) calculated as a full spectrum for each frequency of interest (specific prominence ratio, SPR), both with and without recognition only of pure tones, along with the PR tolerance line.

TNR and PR fail since the corresponding tolerance lines are not exceeded. Only SPR shows a marginal value for a signal with a clearly prominent tonality (even though at a very low sound pressure level).

[34] The calculation of the tone-to-noise Ratio is described in ECMA-418-1.

[35] The calculation of the prominence ratio is described in ECMA-418-1.

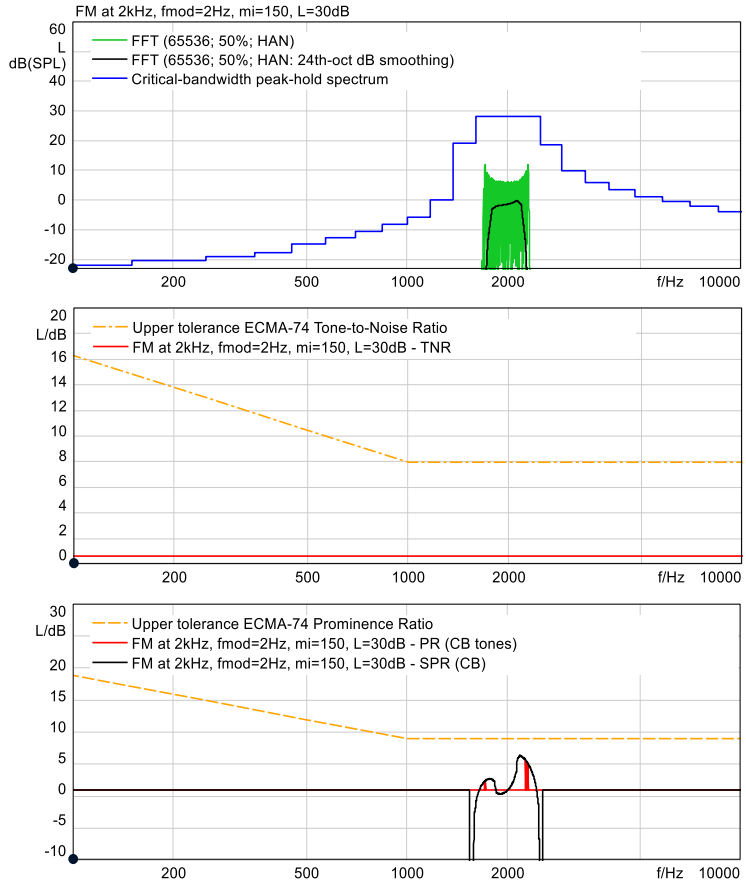

Figure B.1— Top: Different spectral representations (FFT and smoothed FFT) of a frequency-modulated tone; Middle: corresponding TNR results (scale gives dB of tonal audibility); Bottom: corresponding PR values tones-only (according to ECMA-418-1) using critical bands (CB) and complete SPR not constrained only to pure tones results

Figure B.2 depicts the specific psychoacoustic tonality analysis of the same sound as shown in Figure B.1 with a distinct tonal content: The location of the maximum of the specific tonality is changing over time, but the magnitude is almost constant, leading to a stable tonality prediction based on the assumption that the perceived tonality is taken as the maximum of the specific tonality. This corresponds well to the auditory impression.

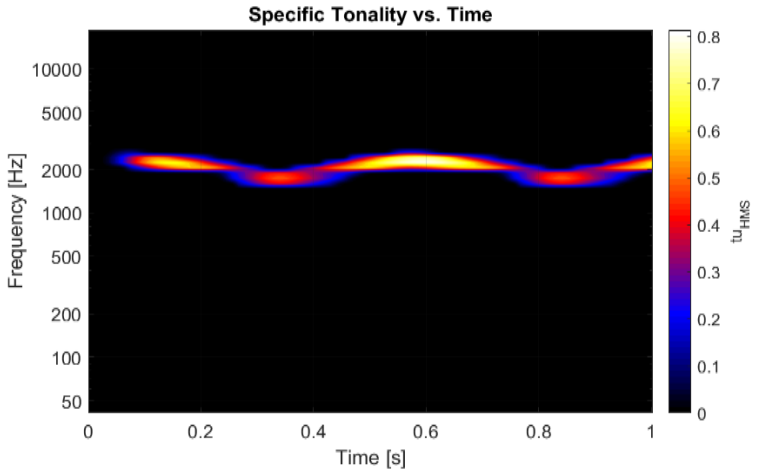

Figure B.2— Specific psychoacoustic tonality analysis of the same sound used as source for the results of the analyses shown in Figure B.1

B.2 Evaluation

The psychoacoustic tonality is evaluated by comparison with listening test results. As a reference, PR is also added to the comparison. TNR values were also calculated. However, since they were very similar to the results of the PR, they are not displayed in the results for reasons of clarity.

For the listening tests, mixtures of a sinusoidal tone with a frequency of 1000 Hz with different levels and pink noise with different levels were used. Thus, the effect of different signal-to-noise-ratios can be evaluated for different levels. Five different tests were performed. In all five tests, the level of the pink noise was varied from 40 dB SPL to 80 dB SPL with a step size of 5 dB SPL. The tests differed in the level of the sinusoidal tone, which was chosen from 55 dB SPL to 75 dB SPL with a step size of 5 dB SPL.

The tests were performed with 16 test subjects. The test subjects were asked to rate the tonality of each sound on a 13-point categorical scale (ranging from “0 - not tonal” to “12 - extremely tonal”). To compare the results of the listening tests with the results of the psychoacoustic model, a linear scaling factor was used for the results of the listening tests. Another scaling factor was used to map the results of the listening tests to the results of the PR. The scaling factors were derived by minimizing the root-mean-square error between the mean ratings of all participants and the calculated psychoacoustic tonality (or the PR, respectively) of all five experiments.

The results of the evaluation are shown in Figure B.3. The results illustrate one problem of the PR: it decreases linearly for decreasing SNR. The tonality perception however does not decrease linearly according to the experimental results. The results of the psychoacoustic hearing model fit much better to the perceived tonality.

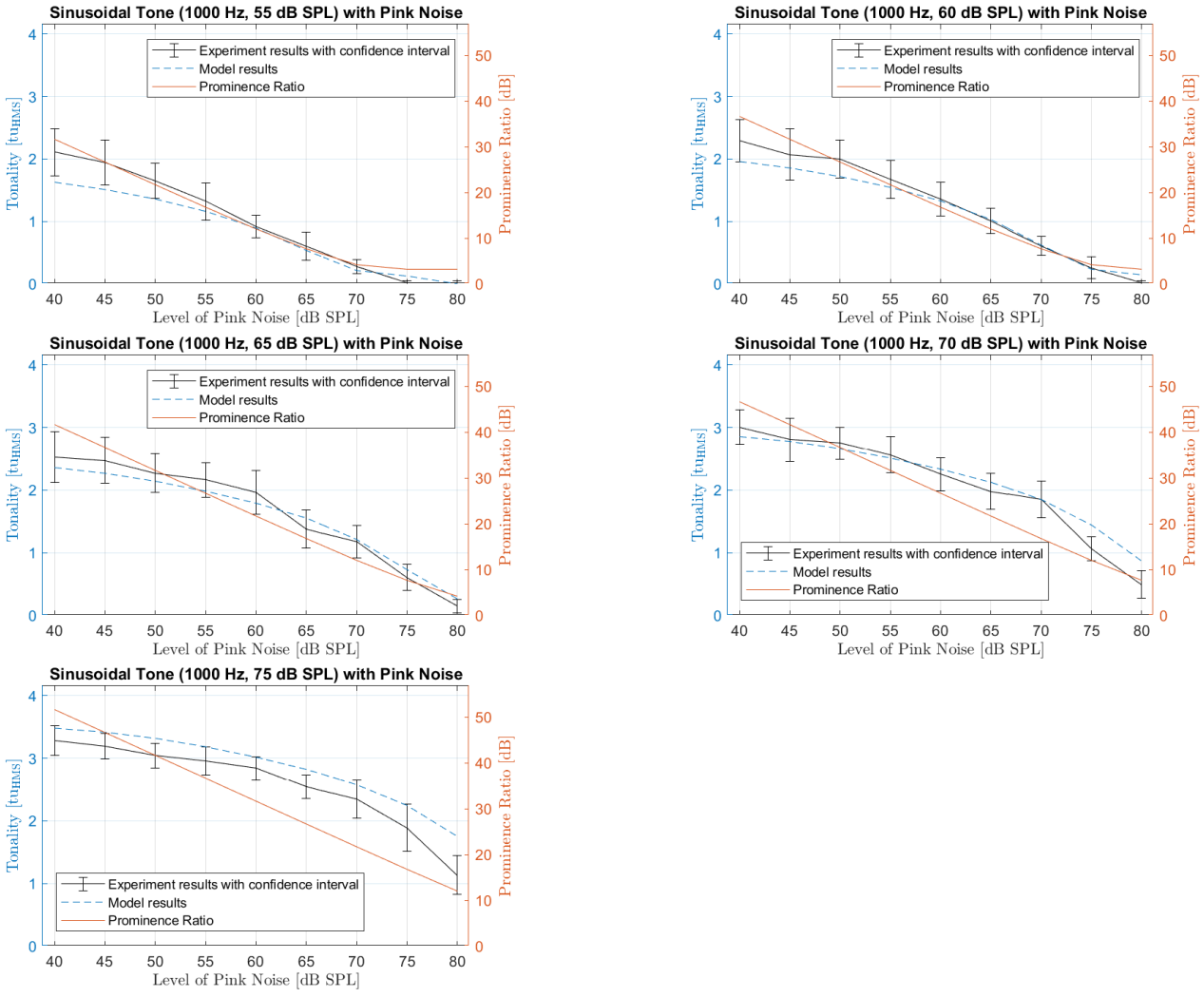

Figure B.3 — Psychoacoustic tonality and prominence ratio compared to results of listening tests

Since experimental results are subject to statistical uncertainty, the variance of the results need to be considered. Thus, an error measure was defined, taking into account the 95% confidence interval of the results. First, the results of the psychoacoustic tonality were scaled such that they are comparable to the tonality ratings of the listening tests. The experimental results are compared to the scaled psychoacoustic tonality. If the psychoacoustic tonality lies within the 95% confidence interval, no error is assumed. If it is outside of the confidence interval, the error is taken as the difference to the confidence interval. The root-mean-square error of this value is calculated. An error for the PR was calculated in the same way, scaling the PR to make it comparable with the results of the listening tests.

The better performance of the psychoacoustic hearing model is also reflected in this error measure. For the psychoacoustic tonality, the error measure over all five experiments (related to the 13-pt categorical scale) was 0,21, for the PR it was 0,70, for the TNR (not shown in the figures) it was 0,74.

Further application examples related to IT equipment can be found in Reference [34].

Annex C
(informative)

Evaluation of the psychoacoustic roughness calculation method

The psychoacoustic roughness is evaluated by comparison with listening test results and data from Reference [12]. Figure C.1 shows results for amplitude modulated sinusoids with seven different carrier frequencies (125 Hz, 250 Hz, 500 Hz, 1000 Hz, 2000 Hz, 4000 Hz, 8000 Hz) and different modulation rates. The results from Reference [12] are idealized, smoothed curves that were fit to the results of jury tests. The results of the model are close to these idealized curves and never exceed a tolerance of ±0,1 asper.

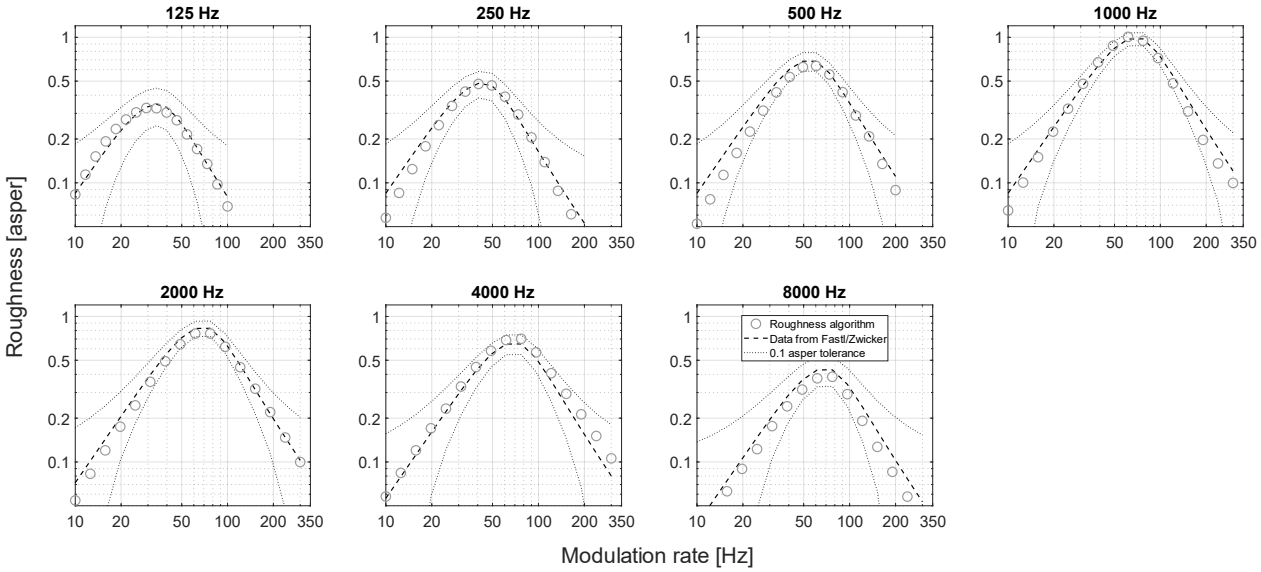

Figure C.1 — Results for modulated sinusoids with different carrier frequencies and modulation rates. All sounds were modulated with 100% degree of modulation and a sound pressure level of 60 dB.

In order to investigate the applicability of the method to technical sounds, listening tests with the sounds described in Table C.1 were carried out.

Table C.1 — Technical Sounds

ESA_02	Electrical Seat Adjuster
ETB_01	Electrical Toothbrush
GEN_02	Generator
HDD_07	Hard Disk Drive
HDD_09	Hard Disk Drive
SCOOT	Pass-by of Scooter
SINUS	Calibration Signal: Modulated Sinus Tone
TOF_03	Take-Off (Airplane)

In Figure C.2, the results of the psychoacoustic roughness model are compared with listening test results (mean values and 95% confidence intervals) for the seven technical sounds and a reference sound (SINUS), which was used as anchor. It can be seen that the calculated results are all within the 95% confidence intervals of the listening test data, thus proving that the algorithm performs well for technical sounds. More results can be found in Reference [30].

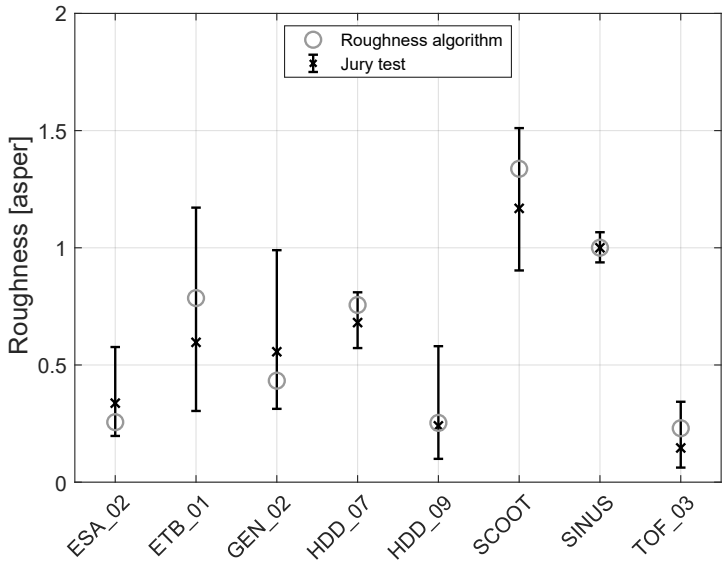

Figure C.2 — Results of several technical sounds. The results of the listening tests are displayed: mean values with 95% confidence intervals.

Bibliography

[1] R. Sottek: A Hearing Model Approach to Time-Varying Loudness, Acta Acustica united with Acustica, vol. 102, no. 4, pp. 725-744, 2016.

[2] M. Slaney: *Auditory toolbox*. Interval Research Corporation, Tech. Rep 10 (1998), 1998.

[3] ISO 532-1: Acoustics – Methods for calculating loudness, Part 1: Zwicker method

[4] 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.

[5] J. Chalupper, H. Fastl: Dynamic loudness model (DLM) for normal and hearing-impaired listeners, Acta Acustica united with Acustica 88(3), pp. 378-386, 2002.

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

[7] ISO 532-2, Acoustics — Methods for calculating loudness — Part 2: Moore-Glasberg method

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

[9] 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, vol. 101(6), pp. 1145–1156, 2015.

[10] ISO 389-7, Acoustics — Reference zero for the calibration of audiometric equipment — Part 7: Reference threshold of hearing under free-field and diffuse-field listening conditions

[11] A. M. H. J. Aertsen and P. I. M. Johannesma: Spectro-Temporal Receptive-Fields of Auditory Neurons in the Grassfrog .1. Characterization of Tonal and Natural Stimuli, Biological Cybernetics, vol. 38, no. 4, pp. 223-234, 1980.

[12] H. Fastl, E. Zwicker: Psychoacoustics. Facts and Models, Springer, Berlin, Heidelberg, New York, 2006.

[13] B. C. Moore: *Basic auditory processes involved in the analysis of speech sounds*. Philosophical Transactions of the Royal Society of London B: Biological Sciences, 363. Jg., Nr. 1493, S. 947-963, 2008.

[14] T. Bierbaums, R. Sottek: *Modellierung der zeitvarianten Lautheit mit einem Gehörmodel*, Proc. DAGA 2012, Darmstadt, pp. 591-592, 2012.

[15] S. Buus, M. Florentine: *Modifications to the power function for loudness*. In: E. Summerfield, R. Kompass, T. Lachmann (eds), Fechner Day 2001. Proceedings of the 17th Annual Meeting of the International Society for Psychophysics. Berlin: Pabst, pp. 236-241, 2001.

[16] M. Epstein, M. Florentine: A test of the Equal-Loudness-Ratio hypothesis using cross-modality matching functions, J. Acoust. Soc. Am., vol. 118(2), pp. 907-913, 2005.

[17] R. Sottek: Improvements in calculating the loudness of time varying sounds. Proc. Inter-Noise 2014, Melbourne, 2014.

[18] R. Sottek: Modelle zur Signalverarbeitung im menschlichen Gehör, dissertation, RWTH Aachen, 1993.

[19] R. Sottek, F. Kamp, A. Fiebig: *A new hearing model approach to tonality*, Proc. Inter-Noise 2013, Innsbruck, 2013.

[20] R. Sottek: *Progress in calculating tonality of technical sounds*, Proc. Inter-Noise 2014, Melbourne, 2014.

[21] R. Sottek: Calculating tonality of IT product sounds using a psychoacoustically-based model, Proc. Inter-Noise 2015, San Francisco, 2015.

[22] H. Hansen, J.L. Verhey, R. Weber: *The Magnitude of Tonal Content*. A Review, Acta Acustica united with Acustica, 97(3), pp. 355-363, 2011.

[23] H. Hansen, R. Weber: Zum Verhältnis von Tonhaltigkeit und der partiellen Lautheit der tonalen Komponenten in Rauschen, Proc. DAGA 2010, Berlin, pp. 597-598, 2010.

[24] J.L. Verhey, S. Stefanowicz: Binaurale Tonhaltigkeit, Proc. DAGA 2011, Düsseldorf, pp. 827-828, 2011.

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

[26] R. Sottek: Gehörgerechte Rauhigkeitsberechnung, Proc. DAGA 1994, Dresden, pp. 1197-1200, 1994.

[27] R. Sottek, P. Vranken, H.-J. Kaiser: Anwendung der gehörgerechten Rauhigkeitsberechnung, Proc. DAGA 1994, Dresden, pp. 1201-1204, 1994.

[28] R. N. Bracewell: The Fourier transform and its applications. McGraw-Hill, New York,1986.

[29] J. Becker, R. Sottek: Psychoacoustic Tonality Analysis, Proc. Inter-Noise 2018, Chicago, 2018.

[30] R. Sottek, J. Becker, T. Lobato: Progress in Roughness Calculation, Proc. Inter-Noise 2020, Seoul, 2020.

[31] A. Oetjen, "Threshold and Suprathreshold Phenomena in Auditory Modulation Perception" (Phd.-Thesis), Oldenburg, 2018.

[32] A. Oetjen, U. Letens, S. van de Par, J. Verhey und R. Weber, „Roughness calculation for randomly modulated sounds," in DAGA, Meran, 2013.

[33] Moore B C, Glasberg B R, Varathanathan A, Schlittenlacher, J. A loudness model for time-varying sounds incorporating binaural inhibition. Trends in hearing, 20, 2331216516682698, 2016.

[34] HEAD acoustics GmbH: Using the new psychoacoustic tonality analyses Tonality (Hearing Model), Application Note, 2018.

[35] Zwicker, E., Loudness and excitation patterns of strongly frequency modulated tones, in Sensation and Measurement, papers in honor of S.S. Stevens, edited by H.R. Moskowitz, B. Scharf, and J.C. Stevens (D. Reidel, Dordrecht, Netherlands), pp. 325–335, 1974.

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

[37] Hots, J. et al., Loudness of sounds with a subcritical bandwidth: A challenge to current loudness models? J. Acoust. Soc. Am. 134(4), EL334–EL339, 2013.

[38] Hots, J. et al., Loudness of subcritical sounds as a function of bandwidth, center frequency, and level. J. Acoust. Soc. Am., 135(3), pp. 1313-1320, 2014.

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