DOI: 10.61782/fa.2025.0575

MODELING FLUCTUATION STRENGTH BASED ON THE SOTTEK HEARING MODEL

Roland Sottek[1*]

[1] HEAD acoustics GmbH, Germany

ABSTRACT

Fluctuating sounds are easily recognizable and have a significant impact on sound quality. It is therefore essential to quantify them in a way that reflects human perception. Extensive research in the literature has been conducted on the perception of fluctuating sounds. However, there is currently no standardized calculation method. There was no reliable approach for estimating the perceived fluctuation strength, especially for technical sounds. This paper presents an algorithm for calculating the perceived fluctuation strength of technical sounds, extending a method previously presented in DAGA 2023. The algorithm is based on the Sottek Hearing Model Roughness published in the ECMA-418-2 standard (3rd edition) and the **HSA** (**H**igh-resolution **S**pectral **A**nalysis) to identify low-rate modulations. It was improved and validated using the results of listening experiments with technical sounds and synthetic data. The algorithm is proposed for inclusion in the 4th edition of the ECMA-418-2 standard.

Keywords: Fluctuation strength, Sottek Hearing Model, High-resolution Spectral Analysis (HSA), ECMA-418-2, psychoacoustics.

1. INTRODUCTION

Temporal variations in sounds easily attract the listener's attention and significantly affect sound quality. Therefore, their proper quantification with respect to human perception is an important task. The auditory sensations roughness and fluctuation strength describe the perception of such temporal variations in sounds. While fluctuation strength covers slow variations (typically below 20 Hz), roughness is produced by faster variations up to about 500 Hz. The maximum of the auditory sensation is located at a modulation rate of about 4 Hz for the fluctuation strength and 70 Hz for the roughness.

Fluctuation strength is used for the perceptual evaluation of sound characteristics as well as for sound design, e.g., for warning sounds. As fluctuation strength increases, sounds become more noticeable and are perceived as increasingly annoying, without any difference in loudness or A-weighted sound pressure level.

Fluctuation strength depends on the modulation rate f_{mod}, the degree of modulation m and the sound pressure level. Frequency modulated sounds produce a similar fluctuation strength as amplitude modulated sounds. Compared to roughness, fluctuation strength is only slightly dependent on the carrier frequency. The unit of fluctuation strength is "vacil$_{\text{HMS}}$". As reference signal with $F = 1$ vacil$_{\text{HMS}}$, an amplitude modulated sinusoid of 1 kHz carrier frequency, $m = 1$, $f_{\text{mod}} = 4$ Hz and a sound pressure level of 60 dB was chosen.

The perception of fluctuating sounds has been widely studied [1-5]. However, there is currently no standardized calculation method. This paper describes an algorithm for calculating the perceived fluctuation strength of synthetic and technical sounds that is proposed for inclusion in the 4th edition of the ECMA-418-2 standard. The validation of the new method is published in [6].

2. SOTTEK HEARING MODEL FLUCTUATION STRENGTH VS. ROUGHNESS

The fluctuation strength algorithm is similar to the roughness calculation based on the Sottek Hearing Model

*Corresponding author: roland.sottek@head-acoustics.com.

as described in the ECMA-418-2 standard [7] using the specific basis loudness $N'_{\text{basis}}(l, z)$ as the starting point for the calculation. The specific basis loudness, depending on the block index l resulting from the segmentation of the bandpass signals (overlap 75%) and the critical band rate scale values z (53 critical band filters with an overlap of 50%), is determined from the original time signal $p(n)$ by the steps shown in Figure 1. This model considers many aspects of auditory perception [7-8], such as the filtering of the outer and middle ear, the auditory filter bank, and the compressive non-linearity of human hearing.

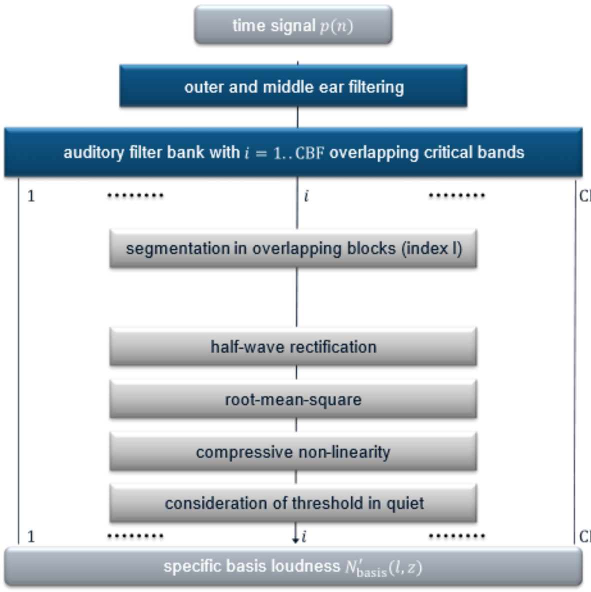

Figure 1. Structure of the Sottek Hearing Model for the calculation of the specific basis loudness, including the auditory filter bank, where CBF is the number of critical band filters in the filter bank.

Figure 2 shows all processing steps for calculating the specific fluctuation strength $F'(l_{50}, z)$, where l_{50} is the block index after interpolation, which starts the last processing step as for the roughness calculation. The fluctuation strength calculation is based on scaled envelope power spectra, which are calculated using the envelope of the segmented critical band signals $p_{l,z}(n')$. The spectral weighting of the envelope spectra in each critical band for the fluctuation strength modeling was adjusted compared to the roughness algorithm to obtain weighting factors compatible with the lower modulation rates of fluctuating sounds.

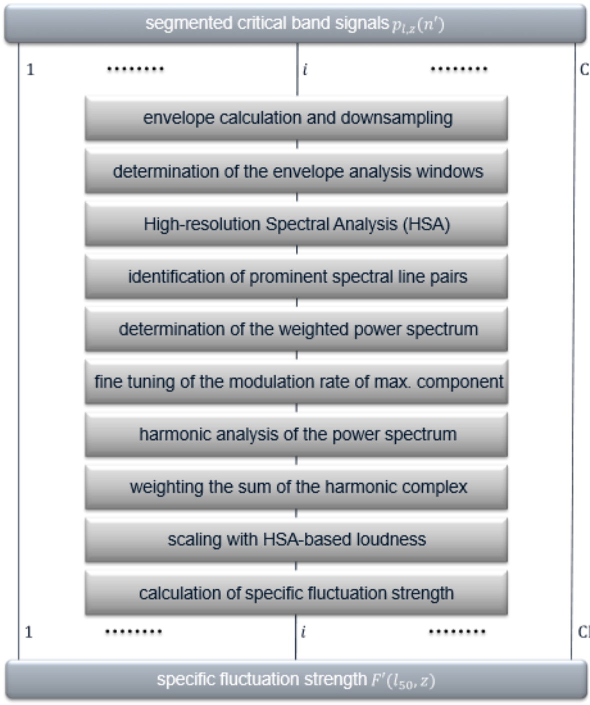

Figure 2. Calculation of specific fluctuation strength based on segmented critical band signals.

In addition, spectral analysis is more challenging at low modulation rates because the constant part of the envelope interferes with the spectral estimation, especially at low modulation rates. To improve the spectral estimation, High-resolution Spectral Analysis (HSA) [9] is introduced along with envelope-dependent analysis windows in order to reduce artifacts due to the envelope calculation and to reduce the influence of quieter periods in the signal. A quieter period within a block of a discrete-time signal is an interval in which all values are below a threshold value and the values to the left and right of this interval are greater than or equal to this threshold value. If the quieter period is at the beginning or end of a block, only the value to the right or left of this interval must be greater than or equal to this threshold value. By using HSA instead of DFT for spectral analysis, the noise reduction step used in the roughness algorithm [7] is eliminated.

The HSA method is an approach to extract periodic components from the signal with very high time and frequency resolution. It works by deconvolution of the original signal spectrum into different sinusoids (with possible interaction between them) and can achieve theoretically infinite resolution for signals without noise and considerably high resolution for signals with noise.

3. FLUCTUATION STRENGTH ALGORITHM

3.1 Envelope calculation and downsampling

The low-frequency envelopes are calculated from the segmented bandpass filtered sound pressure signals $p_{l,z}(n')$ using the Hilbert transform (denoted $\mathcal{H}(\cdot)$). The envelopes $p_{\mathrm{E},l,z}(n')$ are taken as magnitude of the analytic signals

$$p_{\mathrm{E},l,z}(n') = \left|p_{l,z}(n') + j\mathcal{H}\left(p_{l,z}(n')\right)\right|. \quad (1)$$

Since the envelope curves contain only low modulation rates, they are downsampled by a factor of 32, leaving the 1st sample and then every 32nd sample without anti-aliasing low-pass filtering to achieve higher efficiency, which, however, is accompanied by a very slight deviation. The resulting downsampled envelopes of the bandpass signals are denoted $p_{\mathrm{E},l,z}(\tilde{n})$, $\tilde{n}$ refers to the index of the downsampled signal. With this step, the sampling rate changes from $r_s = 48$ kHz to $\tilde{r}_s = 1500$ Hz. The block size $\tilde{s}_\mathrm{b} = 2048$ and a hop size of $\tilde{s}_\mathrm{h} = 512$ are the values corresponding to the block size of $s_\mathrm{b} = 65536$ and the hop size of $s_\mathrm{h} = 16384$ for the segmentation.

3.2 Determination of envelope analysis windows

For the spectral analysis of the envelopes $p_{E,l,z}(\tilde{n})$ only a portion of the samples in each block (size $\tilde{s}_\mathrm{b}$) is considered. Due to the Hilbert transform in Eqn. (1) there may be distortions at the beginning and end of a block. The influence of these distortions shall be reduced by defining an envelope-dependent analysis window for $\tilde{n} = 0, \dots, \tilde{s}_\mathrm{b} - 1$:

$$w_{\mathrm{E},l,z}(\tilde{n}) = \begin{cases} 1, & n_{\mathrm{zb},l,z} \leq \tilde{n} \leq \tilde{s}_\mathrm{b} - 1 - n_{\mathrm{ze},l,z} \\ 0, & \text{else} \end{cases} \quad (2)$$

where $n_{\mathrm{zb},l,z}$.and $n_{\mathrm{ze},l,z}$ correspond to the number of zeros in the analysis window at the beginning and at the end. As a starting point both variables are set to $\tilde{s}_\mathrm{b}/32 = 64$ to reduce possible distortion due to the Hilbert transform. The number of ones in the analysis window then equals $\tilde{n}_\mathrm{ones} = \tilde{s}_\mathrm{b} \cdot 15/16 = 1920$.

The effects of quieter periods shall also be considered. To detect quieter periods, the envelopes $p_{E,l,z}(\tilde{n})$ are smoothed by a moving median filter [1] with a length of $\tilde{s}_\mathrm{b}/64 + 1 = 33$ and then rounded to 8 digits to the right of the decimal point in order to reduce differences due to different implementations. Next, $p_{E,l,z}(\tilde{n})$ is multiplied by $w_{\mathrm{E},l,z}(\tilde{n})$ with the initial parameters of $n_{\mathrm{zb},l,z}$.and $n_{\mathrm{ze},l,z}$, resulting in $\overline{p}_{\mathrm{E},l,z}(\tilde{n})$. Then $\overline{p}_{\mathrm{Emax},l,z}$ is calculated as the maximum of $\overline{p}_{\mathrm{E},l,z}(\tilde{n})$ for $\tilde{n} = 0, \dots, \tilde{s}_\mathrm{b} - 1$.

If $\overline{p}_{\mathrm{Emax},l,z} > 5 \cdot 10^{-6}$ Pa, the detection of quieter periods is continued. Otherwise the entire block is considered to be a quieter period: $w_{\mathrm{E},l,z}(\tilde{n})$ is set to zero for $\tilde{n} = 0, \dots, \tilde{s}_\mathrm{b} - 1$ and $\mathrm{n}_{\mathrm{zb},l,z}$ as well as $n_{\mathrm{ze},l,z}$ are set to $\tilde{s}_\mathrm{b}/2 = 1024$.

A quieter period of $\overline{p}_{\mathrm{E},l,z}(\tilde{n})$ is defined such that the following relationship holds

$$\begin{aligned} &\overline{p}_{\mathrm{E},l,z}(\tilde{n}) < \overline{p}_{\mathrm{Ethr},l,z} \text{ for } \tilde{n} \in \left[\tilde{n}_{\mathrm{qpb},l,z}\ \tilde{n}_{\mathrm{qpe},l,z}\right] \wedge \\ &\overline{p}_{\mathrm{E},l,z}\left(\tilde{n}_{\mathrm{qpb},l,z} - 1\right) \geq \overline{p}_{\mathrm{Ethr},l,z} \wedge \\ &\overline{p}_{\mathrm{E},l,z}\left(\tilde{n}_{\mathrm{qpe},l,z} + 1\right) \geq \overline{p}_{\mathrm{Ethr},l,z} \end{aligned} \quad (3)$$

with $\overline{p}_{\mathrm{Ethr},l,z} = 0.01 \cdot \overline{p}_{\mathrm{Emax},l,z}$, also rounded to 8 digits to the right of the decimal point. There can be several quieter periods, each starting at a different value for $\tilde{n} = \tilde{n}_{\mathrm{qpb},l,z}$ and ending at a different value for $\tilde{n} = \tilde{n}_{\mathrm{qpe},l,z}$.

In the next step, quieter periods at the beginning and the end are determined. The parameters $n_{\mathrm{zb},l,z}$ and $n_{\mathrm{ze},l,z}$ are updated:

$$\begin{aligned} n_{\mathrm{zb},l,z} &= \underset{\tilde{n}=0,\dots,\tilde{s}_\mathrm{b}-1}{\operatorname{argmin}} \left(\overline{p}_{\mathrm{E},l,z}(\tilde{n}) \geq \overline{p}_{\mathrm{Ethr},l,z}\right) \\ n_{\mathrm{ze},l,z} &= \tilde{s}_\mathrm{b} - 1 - \underset{\tilde{n}=0,\dots,\tilde{s}_\mathrm{b}-1}{\operatorname{argmax}} \left(\overline{p}_{\mathrm{E},l,z}(\tilde{n}) \geq \overline{p}_{\mathrm{Ethr},l,z}\right) \end{aligned} \quad (4)$$

Only the quieter period $\left[\tilde{n}_{\mathrm{qpmb},l,z}\ \tilde{n}_{\mathrm{qpme},l,z}\right]$ with the longest duration $\tilde{n}_\mathrm{zeros} = \tilde{n}_{\mathrm{qpe},l,z} - \tilde{n}_{\mathrm{qpb},l,z} + 1$ in the *updated* interval $\tilde{n} \in \left[n_{\mathrm{zb},l,z}\ n_{\mathrm{ze},l,z}\right]$ is determined, where $\tilde{n}_\mathrm{zeros}$ must also be greater than $\tilde{n}_\mathrm{zeros,min} = \tilde{s}_\mathrm{b} \cdot 5/32 = 320$. In the following only the parameters $\tilde{n}_{\mathrm{qpmb},l,z}$ and $\tilde{n}_{\mathrm{qpme},l,z}$ of the longest quieter period are required.

If a valid quieter period $\left[\tilde{n}_{\mathrm{qpmb},l,z}\ \tilde{n}_{\mathrm{qpme},l,z}\right]$ was found in the last step, there are two candidates for the analysis window parameters $n_{\mathrm{zb},l,z}$.and $n_{\mathrm{ze},l,z}$. Depending on the parameters $\tilde{n}_{\mathrm{qpmb},l,z}$ and $\tilde{n}_{\mathrm{qpme},l,z}$ of this quieter period within the entire block under consideration, the part on the left or right that leads to a longer part with ones is used to further update the parameters $n_{\mathrm{zb},l,z}$.and $n_{\mathrm{ze},l,z}$ of $w_{\mathrm{E},l,z}(\tilde{n})$ in Eqn. (2).

If the difference between the beginning of the quieter period $\tilde{n}_{\mathrm{qpmb},l,z}$ and $n_{\mathrm{zb},l,z} + 64$ is greater than the difference between $\tilde{s}_\mathrm{b} - 1 - n_{\mathrm{ze},l,z} - 64$ and the end of the quieter period $\tilde{n}_{\mathrm{qpme},l,z}$, then only $n_{\mathrm{ze},l,z}$ is updated to $\tilde{s}_\mathrm{b} - 1 - \tilde{n}_{\mathrm{qpmb},l,z} + 64$, otherwise only $n_{\mathrm{zb},l,z}$ is updated to

[1] A sliding window is centered about the element in the current position. The window size is automatically truncated at the endpoints when there are not enough elements to fill the window. When the window is truncated, the median is taken over only the elements that fill the window.

$\tilde{n}_{\mathrm{qpme},l,z} + 64$. The updated values also take into account that possible distortions due to the Hilbert transform are reduced by adding $\tilde{s}_{\mathrm{b}}/32 = 64$ zeros.

The final step is to check whether the remaining interval $\tilde{n} \in [\tilde{n}_{1,l,z}\ \tilde{n}_{2,l,z}]$ with $\tilde{n}_{1,l,z} = n_{\mathrm{zb},l,z}$, $\tilde{n}_{2,l,z} = \tilde{s}_{\mathrm{b}} - 1 - n_{\mathrm{ze},l,z}$, contains enough information to estimate the envelope spectrum at lower modulation rates. The first constraint is $\tilde{n}_{2,l,z} - \tilde{n}_{1,l,z} + 1 \geq \tilde{s}_{\mathrm{b}} \cdot 5/32 = 320$ and the second constraint is $\tilde{n}_{2,l,z} \geq \tilde{s}_{\mathrm{b}}/4 - 1 = 511$. The final proof is to calculate the relative standard deviation of a linear regression analysis of $\overline{p}_{\mathrm{E},l,z}(\tilde{n})$ based on a least-squares error minimization for $\tilde{n} = \tilde{n}_{1,l,z} + 10, \dots, \tilde{n}_{2,l,z} - 10$. The relative standard deviation should be at least 0.1%.

If these requirements are not fulfilled, the complete block is considered to be a quieter period: $w_{\mathrm{E},l,z}(\tilde{n})$ is set to zero for $\tilde{n} = 0, \dots, \tilde{s}_{\mathrm{b}} - 1$ and $\mathrm{n}_{\mathrm{zb},l,z}$.as well as $n_{\mathrm{ze},l,z}$ to $\tilde{s}_{\mathrm{b}}/2 = 1024$.

3.3 High-resolution Spectral Analysis (HSA)

The envelopes $p_{\mathrm{E},l,z}(\tilde{n})$ are windowed with $w_{\mathrm{E},l,z}(\tilde{n})$ as described in Eqn. (2) (using the parameters $n_{\mathrm{zb},l,z}$ and $n_{\mathrm{ze},l,z}$ as determined in the last section) and both the corresponding spectrum $\mathrm{P}_{\mathrm{E},l,z}(k)$ and power spectrum $\Phi_{\mathrm{E},l,z}(k)$ are calculated:

$$\begin{aligned} \mathrm{P}_{\mathrm{E},l,z}(k) &= \mathrm{DFT}_{\tilde{s}_{\mathrm{b}}}\left(p_{\mathrm{E},l,z}(\tilde{n}) \cdot w_{\mathrm{E},l,z}(\tilde{n})\right) \\ \Phi_{\mathrm{E},l,z}(k) &= \left|\mathrm{P}_{\mathrm{E},l,z}(k)\right|^2 \end{aligned} \tag{5}$$

where $\mathrm{DFT}_{\tilde{s}_{\mathrm{b}}}$ denotes the $\tilde{s}_{\mathrm{b}}$-point Discrete Fourier Transform [2], k is the index corresponding to a modulation rate of $k \cdot \Delta f$ with $\Delta f = \frac{\tilde{r}_S}{\tilde{s}_{\mathrm{b}}}$ Hz.

Only values for $k = 0, \dots, 48$ are used to predict the result of the HSA, the vector $\widehat{\mathbf{P}}_{\mathrm{E,HSA},l,z}$ consisting of a constant part (modulation rate equal to 0) and M_{c} spectral lines at an arbitrary modulation rate [3] $\mathbf{f}_{\mathrm{c}} = (0, f_{\mathrm{c},1}, \dots, f_{\mathrm{c},M_{\mathrm{c}}})$, (not limited to the discrete resolution of the DFT). The corresponding conjugate complex spectral lines [4] at the corresponding negative frequencies ($-f_{\mathrm{c},m}$, e.g., for the mth complex line) are considered for the mathematical derivation of the following equations, but are omitted in the result, as they provide no additional information.

$\widehat{\mathbf{P}}_{\mathrm{E,HSA},l,z}$ represents the spectrum predicted by the HSA with the influence of the analysis window removed, thus providing not only a result at an arbitrary modulation rate, but also with amplitudes at very high resolution. This is a major advantage of the HSA over standard spectral analysis such as DFT, and also provides more accurate and useful results for other spectral analysis applications. It works by deconvolution of the original signal spectrum into different sinusoids (with possible interaction between them) and can achieve theoretically infinite resolution for signals without noise and considerably high resolution for signals with noise.

The result at the modulation rate 0 is real, the other spectral lines are complex and described by their real and imaginary parts [5]. To simplify the notation, the variable $\mathbf{x}$ is used instead of $\widehat{\mathbf{P}}_{\mathrm{E,HSA},l,z}$: $\mathbf{x} = (x_1, \dots, x_{2 \cdot M_c + 1})^T$ with the elements

$$x_i = \begin{cases} \hat{p}_{0,l,z}, & i = 1 \\ \frac{\Re(\hat{p}_{f_{\mathrm{c},m},l,z})}{2}, & \mathrm{mod}(i,2) = 0 \\ \frac{\Im(\hat{p}_{f_{\mathrm{c},m},l,z})}{2}, & \mathrm{mod}(i,2) \neq 0 \wedge i > 1 \end{cases} \tag{6}$$

where $\hat{p}_{0,l,z}$ is the real constant, $\hat{p}_{f_{\mathrm{c},m},l,z}$ is the mth complex line ($m = 1, \dots, M_{\mathrm{c}}$) and $\mathrm{mod}(i, n)$ returns the remainder after dividing i by n, where i is the dividend and n is the divisor (modulo operation).

The vector $\mathbf{x}$ is determined for a given set of modulation rates $\mathbf{f}_{\mathrm{c}}$ in order to obtain the smallest error

$$E_{l,z}(\mathbf{f}_{\mathrm{c}}) = \sum_{k=0}^{K_{\mathrm{L}}-1} \left|\widehat{\mathrm{P}}_{\mathrm{E},l,z,\mathbf{f}_{\mathrm{c}}}(k) - \mathrm{P}_{\mathrm{E},l,z}(k)\right|^2, \tag{7}$$

with [6]

$$K_{\mathrm{L}} = \min\left(\max\left(17, \mathrm{round}\left(\max(f_c/\Delta f)\right) + 8\right), 49\right). \tag{8}$$

The spectrum $\widehat{\mathrm{P}}_{\mathrm{E},l,z,f_{\mathrm{c}}}(k)$ in Eqn. (7) corresponds to $\widehat{\mathbf{P}}_{\mathrm{E,HSA},l,z}$ including the influence of the analysis window. [7]

For the calculation of $\widehat{\mathbf{P}}_{\mathrm{E,HSA},l,z}$, the matrix $\mathbf{W} = (W_1, \dots, W_{2 \cdot M_c + 1})$ is required with the elements

$$W_{ik} = \begin{cases} W_{\mathrm{E},l,z,0}(k), & i = 1 \\ W_{\mathrm{E},l,z,f_{\mathrm{c},m}^{+}}(k), & i = 2 \cdot m \\ W^{*}_{\mathrm{E},l,z,f_{\mathrm{c},m}^{-}}(k), & i = 2 \cdot m + 1 \end{cases} \tag{9}$$

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

[3] Here $\mathbf{f}_{\mathrm{c}}$ denotes the vector of modulation rates of the candidates for a spectral line pair under consideration.

[4] The complex conjugate of a complex variable z is denoted as z^*.

[5] $\Re(z)$ is the real part of the complex valued variable z and $\Im(z)$ is the imaginary part.

[6] The round function rounds to the nearest integer.

[7] The calculation of $\widehat{\mathrm{P}}_{\mathrm{E},l,z,\mathbf{f}_{\mathrm{c}}}(k)$ is only necessary for deriving the algorithm.

that can be calculated using the following equation

$$W_{\mathrm{E},l,z,f_{\mathrm{c},m}}(k) = \mathrm{e}^{-j2\pi f_{\mathrm{n}}(k)\cdot(\tilde{s}_{\mathrm{b}}-n_{\mathrm{ze},l,z}+n_{\mathrm{zb},l,z}-1)} \cdot \frac{\sin\left(\pi f_{\mathrm{n}}(k)\cdot(\tilde{s}_{\mathrm{b}}-n_{\mathrm{ze},l,z}-n_{\mathrm{zb},l,z})\right)}{\sin(\pi f_{\mathrm{n}}(k))} \quad (10)$$

where $f_{\mathrm{n}}(k) = \frac{k}{\tilde{s}_{\mathrm{b}}} - \frac{f_{\mathrm{c},m}}{\tilde{r}_s} + \varepsilon_0$ is a normalized frequency; ε_0, the smallest positive computer number so that $1 + \varepsilon_0 > 1$, is added to avoid division by zero, $W_{\mathrm{E},l,z,0}(k)$ is calculated using Eqn. (10) with $f_{\mathrm{c},m} = 0$ and

$$\begin{aligned} W_{\mathrm{E},l,z,f_{\mathrm{c},m}^{+}}(k) &= W_{\mathrm{E},l,z,f_{\mathrm{c},m}}(k) + W_{\mathrm{E},l,z,-f_{\mathrm{c},m}}(k) \\ W_{\mathrm{E},l,z,f_{\mathrm{c},m}^{-}}(k) &= W_{\mathrm{E},l,z,f_{\mathrm{c},m}}(k) - W_{\mathrm{E},l,z,-f_{\mathrm{c},m}}(k) \end{aligned} \quad (11)$$

The optimal values of $\widehat{\mathbf{P}}_{\mathrm{E,HSA},l,z}$ can be calculated by solving a system of $2 \cdot M_{\mathrm{c}} + 1$ equations. This results in a matrix equation of the type $\mathbf{A} \cdot \mathbf{x} = \mathbf{b}$, with a symmetric matrix $\mathbf{A} = (a_{ij})$ and a vector $\mathbf{x}$ as described in Eqn. (6):

$$a_{ij} = \sum_{k=0}^{K_{\mathrm{L}}-1} \Re(W_{\mathrm{ik}}) \cdot \Re(W_{\mathrm{jk}}) + \Im(W_{\mathrm{ik}}) \cdot \Im(W_{\mathrm{jk}}) \quad (12)$$

for $(i, j) \in IJ$, else:

$$a_{ij} = \sum_{k=0}^{K_{\mathrm{L}}-1} \Im(W_{\mathrm{ik}}) \cdot \Re(W_{\mathrm{jk}}) + \Re(W_{\mathrm{ik}}) \cdot \Im(W_{\mathrm{jk}}) \quad (13)$$

using the elements of $\boldsymbol{W}$ according to Eqn. (9) and the following set of indices defined for $i = 1, \ldots, 2 \cdot M_{\mathrm{c}} + 1$ and $j = i, \ldots, 2 \cdot M_{\mathrm{c}} + 1$:

$$\begin{aligned} I_{\mathrm{R}} &= \{i \mid i = 1 \lor \mathrm{mod}(i, 2) = 0\} \\ J_{\mathrm{R}} &= \{j \mid j = 1 \lor \mathrm{mod}(j, 2) = 0\} \\ IJ &= \{(i, j) \mid (i \in I_{\mathrm{R}} \land \mathrm{j} \in J_{\mathrm{R}}) \lor (i \notin I_{\mathrm{R}} \land \mathrm{j} \notin J_{\mathrm{R}})\} \end{aligned} \quad (14)$$

The elements of the vector $\mathbf{b}$ can be calculated as

$$b_i = \sum_{k=0}^{K_{\mathrm{L}}-1} \Re(\mathrm{P}_{\mathrm{E},l,z,k}) \cdot \Re(W_{\mathrm{ik}}) + \Im(\mathrm{P}_{\mathrm{E},l,z,k}) \cdot \Im(W_{\mathrm{ik}}) \quad (15)$$

for $i \in I_R$, else

$$b_i = \sum_{k=0}^{K_{\mathrm{L}}-1} \Im(\mathrm{P}_{\mathrm{E},l,z,k}) \cdot \Re(W_{\mathrm{ik}}) + \Re(\mathrm{P}_{\mathrm{E},l,z,k}) \cdot \Im(W_{\mathrm{ik}}) \quad (16)$$

with $\mathrm{P}_{\mathrm{E},l,z,k} = \mathrm{P}_{\mathrm{E},l,z}(k)$.

The error function $E_{l,z}(\mathbf{f}_{\mathrm{c}})$ can be expressed advantageously with the already calculated coefficients a_{ij}, b_i and the elements x_i of the solution of the system of equations:

$$\begin{aligned} E_{l,z}(\mathbf{f}_{\mathrm{c}}) = & \sum_{k=0}^{K_{\mathrm{L}}-1} \Phi_{\mathrm{E},l,z}(k) + \sum_{i=1}^{2 \cdot M_{\mathrm{c}}+1} a_{ii} \cdot {x_i}^2 + \\ & 2 \cdot \sum_{i=1}^{2 \cdot M_{\mathrm{c}}} \sum_{j=i+1}^{2 \cdot M_{\mathrm{c}}+1} a_{ij} \cdot x_i \cdot x_j - \\ & 2 \cdot \sum_{i=1}^{2 \cdot M_{\mathrm{c}}+1} b_i \cdot x_i \end{aligned} \quad (17)$$

The calculation of the error function $E_{l,z}(\mathbf{f}_{\mathrm{c}})$ is necessary for a further step of the HSA, the fine tuning of the modulation rates in an optimization process as described later. In this case, only one spectral line pair is considered.

3.4 Identification of prominent spectral line pairs

In the next two steps prominent spectral line pairs are identified searching for two cases.

3.4.1 Local maxima of the power spectrum

Local maxima $\Phi_{\mathrm{E},l,z}(k_{\mathrm{p},i}(l, z))$ of the power spectrum $\Phi_{\mathrm{E},l,z}(k)$ fulfilling the condition

$$\Phi_{\mathrm{E},l,z}(k_{\mathrm{p},i}(l, z)) \geq \max(0.001 \cdot \Phi_{\mathrm{E},l,z}(0), \Phi_{\mathrm{Emin}}) \quad (18)$$

with $\Phi_{\mathrm{Emin}} = 0.15$ and $i = 1, \ldots, I_m$. Here, the number of local maxima I_m cannot exceed 24 due to the limited number of spectral lines considered (48 for positive modulation rates). The modulation rates of the remaining local maxima are predicted as

$$f_{\mathrm{p},i}(l, z) = \left(\frac{\sum_{j=-1}^{1}(k_{\mathrm{p},i}(l,z)+j) \cdot \Phi_{\mathrm{E},l,z}(k_{\mathrm{p},i}(l,z)+j)}{\sum_{j=-1}^{1} \Phi_{\mathrm{E},l,z}(k_{\mathrm{p},i}(l,z)+j)} - 1\right) \cdot \Delta f. \quad (19)$$

3.4.2 Local minima of an error function

Local minima of an error function $E_{l,z}((0, f_i))$ according to Eqn. (17) are calculated for modulation rates $f_i = 0.25 \cdot 2^{(i-2)/3}$ Hz, with $i = 1, \ldots, 16$, setting the actual frequency $f_{\mathrm{c},1} = f_i$ of the only candidate of the prominent spectral line pair ($\mathbf{f}_{\mathrm{c}} = (0, f_i)$). At the end, only the component with the modulation rate $f_{\min}(l, z)$ corresponding to the minimum of all local minima of $E_{l,z}((0, f_i))$ is taken as a candidate for a prominent spectral component in addition to the local maxima $\Phi_{\mathrm{E},l,z}(k_{\mathrm{p},i}(l, z))$ at the modulation rates $f_{\mathrm{p},i}(l, z)$. All modulation rates of the candidates for a prominent spectral component are set as elements of the vector $\mathbf{f}_{\mathrm{c}}$ in ascending order, starting with the modulation rate 0.

If there is no local minimum $f_{\min}(l, z)$, all local maxima are considered as candidates for prominent spectral line pairs and the spectrum $\widehat{\mathbf{P}}_{\mathrm{E},l,z,\mathrm{HSA}}$ is obtained using the HSA, providing the constant part and all spectral line pairs at the modulation rates $\mathbf{f}_{\mathrm{c}}$ and the corresponding error $E_{l,z}(\mathbf{f}_{\mathrm{c}})$. Except for the case of no local maximum ($I_m = 0$), there is no modulation in this block ($\widehat{\mathbf{P}}_{\mathrm{E},l,z,\mathrm{HSA}}$ and $E_{l,z}(\mathbf{f}_{\mathrm{c}})$ is set to zero) and further processing for this block is stopped.

3.4.3 Selection of optimal spectral line pairs

If a local minimum $f_{\min}(l, z)$ exists, it is further necessary to check whether there are duplicates of spectral line pairs among the local maxima $\Phi_{\mathrm{E},l,z}(k_{\mathrm{p},i}(l,z))$ at the modulation rates $f_{\mathrm{p},i}(l,z)$. The set

$$I_{i_d} = \left\{ i \,\middle|\, \left| f_{\min}(l,z) - f_{\mathrm{p},i}(l,z) \right| < 1.25 \cdot \Delta f \right\} \tag{20}$$

of the indices of duplicates of spectral line pairs may or may not be empty. [8]
If I_{i_d} is not empty, two cases must be checked to remove duplicates of spectral line pairs:

I. consider $f_{\min}(l,z)$ and all values of $f_{\mathrm{p},i}(l,z)$ for all indices $i = 1, \dots, I_m$ except for the indices in I_{i_d} and calculate the spectrum $\widehat{\mathbf{P}}_{\mathrm{E},l,z,\mathrm{HSA}}$ using the HSA, providing the constant part and all spectral line pairs at the modulation rates $\mathbf{f}_\mathrm{c}$ and the corresponding error $E_{l,z}(\mathbf{f}_\mathrm{c})$,

II. consider only all values of $f_{\mathrm{p},i}(l,z)$ for all indices $i = 1, \dots, I_m$ and calculate the spectrum $\widehat{\mathbf{P}}_{\mathrm{E},l,z,\mathrm{HSA}}$ and the corresponding error $E_{l,z}(\mathbf{f}_\mathrm{c})$.

From the two cases choose the one that the gives the lower error $E_{l,z}(\mathbf{f}_\mathrm{c})$.
If I_{i_d} is empty, then consider $f_{\min}(l,z)$ and all values of $f_{\mathrm{p},i}(l,z)$ for all indices $i = 1, \dots, I_m$ and calculate the spectrum $\widehat{\mathbf{P}}_{\mathrm{E},l,z,\mathrm{HSA}}$ using the HSA, providing the constant part and all spectral line pairs at the modulation rates $\mathbf{f}_\mathrm{c}$ and the corresponding error $E_{l,z}(\mathbf{f}_\mathrm{c})$.
In the next step, only spectral line pairs with a modulation rate satisfying the condition

$$A_i(l,z) > 0.05 \cdot \max_i \left(A_i(l,z) \right) \tag{21}$$

are considered, where $A_i(l,z)$ is the ith preselected component of $\left| \widehat{\mathbf{P}}_{\mathrm{E},l,z,\mathrm{HSA}}(f_{\mathrm{c},i}(l,z)) \right|^2$ and $f_{\mathrm{c},i}(l,z)$ is the modulation rate of the ith preselected spectral line pair, resulting in the subset $\tilde{f}_{\mathrm{c},i}(l,z)$.

3.5 Determination of the weighted power spectrum

Then the weighted power spectrum of the remaining spectral line pairs

$$\tilde{A}_i(l,z) = \left| \widehat{\mathbf{P}}_{\mathrm{E},l,z,\mathrm{HSA}}(\tilde{f}_{\mathrm{c},i}(l,z)) \right|^2 \cdot w_{lh}(\tilde{f}_{\mathrm{c},i}(l,z)) \tag{22}$$

and the modulation rate $\tilde{f}_{\mathrm{c},i_{\max}}(l,z)$ of the maximum of the weighted power spectrum are determined: $i_{\max} = \operatorname{argmax}_i \left(\tilde{A}_i(l,z) \right)$; $w_{lh}(\tilde{f}_{\mathrm{c},i}(l,z))$ weights the power spectrum in relation to lower and higher modulation rates (bandpass characteristic of fluctuation strength): $w_{lh}(0) = 0$,

$$w_{lh}(\tilde{f}_{\mathrm{c},i}(l,z)) = \frac{1}{\left(1 + \left(\left(\frac{\tilde{f}_{c,i}(l,z)}{f_{\max}} - \frac{f_{\max}}{\tilde{f}_{c,i}(l,z)} \right) \cdot q_{1,1} \right)^2 \right)^{q_{2,1}}} \tag{23}$$

for $\tilde{f}_{\mathrm{c},i}(l,z) \leq f_{\max}(z)$, $q_{1,1} = 0.33048$, $q_{2,1} = 0.85902$,

$$w_{lh}(\tilde{f}_{\mathrm{c},i}(l,z)) = \frac{\left(1 + 0.092623 \left| \log_2 \left(\frac{F(z)}{1\,\mathrm{kHz}} \right) \right|^{1.24} \right)^{-1}}{\left(1 + \left(\left(\frac{\tilde{f}_{c,i}(l,z)}{f_{\max}} - \frac{f_{\max}}{\tilde{f}_{c,i}(l,z)} \right) \cdot q_{1,\mathrm{h}} \right)^2 \right)^{q_{2,\mathrm{h}}}} \tag{24}$$

for $\tilde{f}_{\mathrm{c},i}(l,z) \leq f_{\max}(z)$, $q_{1,\mathrm{h}} = 0.21792$, $q_{2,\mathrm{h}} = 4.6728$, $f_{\max} = 4.8659\,\mathrm{Hz}$ is the modulation rate at which the weighting factor reaches the maximum of one and $F(z)$ is the center frequency of the auditory filter bank as described in [7].

3.6 Fine tuning of the modulation rate of the maximum component

The modulation rate of the maximum of the weighted power spectrum $\tilde{f}_{\mathrm{c},i_{\max}}(l,z)$ is used as a starting point x_0 for a fine tuning based on the constant part and only one spectral line pair. The modulation rate $f_{\mathrm{c},1,\mathrm{opt}}(l,z)$, giving the minimum error $E_{l,z}(\mathbf{f}_\mathrm{c})$, is determined by a modified damped Newton method applied to $E'_{l,z}(\mathbf{f}_\mathrm{c})$, the 1st derivative of $E_{l,z}(\mathbf{f}_\mathrm{c})$ with $\mathbf{f}_\mathrm{c} = (0, x_k)$. $E''_{l,z}(\mathbf{f}_\mathrm{c})$, the 2nd derivative of $E_{l,z}(\mathbf{f}_\mathrm{c})$ is also used. Both derivatives are approximated by differential quotients with $\Delta x = 10^{-5}$, the 1st derivative as

$$E'_{l,z}(x_{k-1}) = \frac{\left(E_{l,z}(x_{k-1} + \Delta x) - E_{l,z}(x_{k-1} - \Delta x) \right)}{2 \cdot \Delta x} \tag{25}$$

and the 2nd derivative as

$$E''_{l,z}(x_{k-1}) = \frac{\left(E_{l,z}(x_{k-1} + \Delta x) - 2 \cdot E_{l,z}(x_{k-1}) + E_{l,z}(x_{k-1} - \Delta x) \right)}{(\Delta x)^2} \tag{26}$$

The iteration is started for $k = 1$ and $x_0 = \tilde{f}_{\mathrm{c},i_{\max}}(l,z)$

$$x_k = x_{k-1} - \Delta x_{k-1} \tag{27}$$

with [9]

$$\Delta x_{k-1} = \frac{sign(E'_{l,z}(x_{k-1}))}{4} \cdot \min \left(\frac{|E'_{l,z}(x_{k-1})|}{|E''_{l,z}(x_{k-1})| + \varepsilon_0}, 2 \cdot 10^{-4} \right) \tag{28}$$

[8] Duplicates of spectral line pairs can only occur at slightly higher modulation rates in the given range for $f_{\min}(l,z)$ (here: $f_2 = 0.25\,\mathrm{Hz} \leq f_{\min}(l,z) \leq f_{15} = 0.25 \cdot 2^{\frac{13}{3}} \approx 5.0397\,\mathrm{Hz}$), where also local maxima in the spectrum can be observed.

[9] The signum function is defined as $\mathrm{sign}(x) = \begin{cases} 1, & x > 0 \\ 0, & x = 0. \\ -1, & x < 0 \end{cases}$

and continued until $|\Delta x_{k-1}| > 10^{-7} \wedge k < 40$.

If $|f_{\mathrm{c,1,opt}} - \tilde{f}_{\mathrm{c},i_{\max}}(l,z)| > 1.25 \cdot \Delta f$, it is assumed that the optimization failed and the fine tuning is cancelled, otherwise $\tilde{f}_{\mathrm{c},i_{\max}}(l,z)$ is replaced by $f_{\mathrm{c,1,opt}}(l,z)$ and the corresponding spectral component of $\widehat{\mathbf{P}}_{\mathrm{E},l,z,\mathrm{HSA}}$ as well as $\tilde{A}_{i_{\max}}(l,z)$ are updated accordingly.
If $f_{\mathrm{c,1,opt}} < 0.125$ Hz, then the modulation in this block is discarded ($\widehat{\mathbf{P}}_{\mathrm{E},l,z,\mathrm{HSA}}$ and $E_{l,z}(\mathbf{f}_{\mathrm{c}})$ are set to zero) and further processing for this block is stopped.

3.7 Harmonic analysis of the power spectrum

In this step, the weighted power spectra $\tilde{A}_i(l,z)$ are further analyzed. It is assumed that there is a dominant harmonic complex (a fundamental modulation rate with harmonics at multiples of the fundamental modulation rate) which is the dominant cause of the fluctuation perception. The fundamental modulation rate of such a harmonic complex is estimated in this step. It is further assumed that $\tilde{f}_{\mathrm{c},i_{\max}}(l,z)$ or more precisely the tuned value $f_{\mathrm{c,1,opt}}(l,z)$ is part of the harmonic complex and that the maximum order of this component is three, resulting in three cases to be tested: $\tilde{f}_{\mathrm{c},i_{\max}}(l,z)$ is the fundamental modulation rate or 2nd or 3rd order. Additionally, the highest order to be considered is the 5th order.
For each block l and band z, the fundamental modulation rate of the envelope is estimated in the next processing step, taking into account the modulation rates $\tilde{f}_{\mathrm{c},i}(l,z)$ and the amplitudes $\tilde{A}_i(l,z)$ of the block.
For each assumed order $o = 1,2,3$ of $f_{\mathrm{c,1,opt}}(l,z)$, it is tested whether the corresponding modulation rate $f_{\mathrm{c},1,o}(l,z) = f_{\mathrm{c,1,opt}}(l,z)/o$ is the best estimate for the fundamental modulation rate of the envelope, assuming that the sum over the harmonic complex corresponding to the best estimate gives the highest value. The exact procedure for each assumed order o is described below.
Initially, the integer ratios of all the modulation rates $\tilde{f}_{\mathrm{c},i}(l,z)$ to the modulation rate $f_{\mathrm{c},1,o}(l,z)$ are calculated

$$R_{i,o}(l,z) \quad = \mathrm{round}\left(\frac{\tilde{f}_{\mathrm{c},i}(l,z)}{f_{\mathrm{c},1,o}(l,z)}\right), \tag{29}$$

by rounding to the nearest integer. All integer ratios greater than five are set to zero.
From all remaining values, a set $I_{i,o}$ of indices of all components belonging to a harmonic complex with fundamental modulation rate $f_{\mathrm{c},1,o}(l,z)$ is defined (using a tolerance of 4%):

$$I_{i,o}(l,z) = \left\{ i \,\middle|\, \left(\left| \frac{\tilde{f}_{\mathrm{c},i}(l,z)}{R_{i,o}(l,z) \cdot f_{\mathrm{c},1,o}(l,z)} - 1 \right| < 0.04 \right) \right\}. \tag{30}$$

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

$$E_{i,o}(l,z) \quad = \sum_{i \in I_{i,o}} \tilde{A}_i(l,z). \tag{31}$$

The order o leading to the highest energy is denoted in the following as $o_{\max}$, the corresponding set of indices $I_{i,o}(l,z)$ is denoted as $I_{\max}(l,z)$. The fundamental modulation rate of the envelope is $f_1(l,z) = \tilde{f}_{\mathrm{c},o_{\max}}(l,z)$.
In the following, only the components corresponding to the indices in $I_{\max}(l,z)$ are considered as part of the envelope. The modulation rates of these remaining components, except for $f_{\mathrm{c},1,o}(l,z)$, are corrected according to their integer ratios to $\tilde{f}_{\mathrm{c},o_{\max}}(l,z)$ and the corresponding spectral line pairs of $\widehat{\mathbf{P}}_{\mathrm{E},l,z,\mathrm{HSA}}$ are calculated for the given orders (based on the integer ratios, maximum up to the 5th order) using the HSA with the constant part and one spectral line pair for each order of interest. The constant part is predicted several times for all orders. The mean value of all predictions is taken as the corresponding final result. The weighted power spectra $\tilde{A}_i(l,z)$ are updated for the improved modulation rates.

3.8 Weighting the sum of the harmonic complex

The sum of the harmonic complex is weighted according to the distance between the center of gravity of its components and the modulation rate of the component with the largest amplitude:

$$\hat{A}(l,z) = w_{\mathrm{bw}} \cdot \sum_{i \in I_{\max}(l,z)} \tilde{A}_i(l,z) \tag{32}$$

with

$$w_{\mathrm{bw}} = 1 + c \cdot \left| \frac{\sum_{i \in I_{\max}(l,z)} \left(\frac{\tilde{f}_{\mathrm{c},i}(l,z)}{\mathrm{Hz}} \cdot \tilde{A}_i(l,z) \right)}{\left(\sum_{i \in I_{\max}(l,z)} \tilde{A}_i(l,z) \right) + \varepsilon_0} - \frac{f_{\mathrm{c,1,opt}}(l,z)}{\mathrm{Hz}} \right|^{e} \tag{33}$$

where $c = 0.79577$ and $e = 0.43461$.

3.9 Scaling with HSA-based loudness

Finally, $\hat{A}(l,z)$ is weighted with a factor depending on the power of the harmonic complex and the specific loudness $N'_{\mathrm{HSA}}(l,z)$ based on the HSA:

$$A(l,z) = \frac{\tilde{s}_{\mathrm{b}} \cdot \hat{A}(l,z)}{\hat{p}_{0,l,z}{}^{2} + 2 \cdot \sum_{i \in I_{\max}(l,z)} A_i(l,z) + \varepsilon_0} \, \frac{\left(N'_{\mathrm{HSA}}(l,z) \right)^2 \cdot \left(\frac{\mathrm{Bark}_{\mathrm{HMS}}}{\mathrm{sone}_{\mathrm{HMS}}} \right)}{\max_z \left(N'_{\mathrm{HSA}}(l,z) \right) + \varepsilon_0} \tag{34}$$

$N'_{\mathrm{HSA}}(l,z)$ is calculated by applying the nonlinearity described in [7, section 5.1.8] to the RMS value of the harmonic complex:

$$\tilde{N}'_{\mathrm{HSA}}(l,z) = A'\left(\sqrt{\frac{\hat{p}_{0,l,z}{}^{2} + 2 \cdot \sum_{i \in I_{\max}(l,z)} A_i(l,z)}{2}}\right) \tag{35}$$

and taking into account the lower threshold of hearing $\mathrm{LTQ}(z)$ according to [7, section 5.1.9] by subtraction, whereby negative values are set to zero:

$$N'_{\mathrm{HSA}}(l,z) = \max\left(\tilde{N}'_{\mathrm{HSA}}(l,z) - \mathrm{LTQ}(z), 0\right). \tag{36}$$

All values of $A(l,z)$ below a threshold of 5.2519 are set to zero. The corresponding fundamental modulation rates $f_1(l,z)$ are also set to zero.

3.10 Calculation of specific fluctuation strength

The calculation of the time-dependent specific fluctuation strength is mainly performed as for the roughness, starting with the interpolation of $A(l,z)$ using a piecewise cubic Hermitian function to a sampling rate of $r_{s50} = 50\,\mathrm{Hz}$ as described in [7] resulting in an uncalibrated estimate of the specific fluctuation strength $F'_{\mathrm{est}}(l_{50},z)$. The next step in the calculation of the specific fluctuation strength is a nonlinear transformation, depending on the distribution of $F'_{\mathrm{est}}(l_{50},z)$ over the critical bands z and a calibration:

$$\hat{F}'(l_{50},z) = c_{\mathrm{F}} \cdot \left(F'_{\mathrm{est}}(l_{50},z)\right)^{E(l_{50})} \tag{37}$$

with the calibration factor $c_{\mathrm{F}} = 0.003840572\ \frac{\mathrm{vacil}_{\mathrm{HMS}}}{\mathrm{Bark}_{\mathrm{HMS}}}$,

$$E(l_{50}) = 0.37106 \cdot \frac{\left(\tanh\left(1.6407 \cdot (B(l_{50}) - 2.5804)\right) + 1\right)}{2} + 0.58449 \tag{38}$$

and $B(l_{50})$ as described in the following

$$\hat{B}(l_{50}) = \frac{\tilde{F}'_{\mathrm{est}}(l_{50})}{\bar{F}'_{\mathrm{est}}(l_{50}) + 10^{-12}} \tag{39}$$

where $\tilde{F}'_{\mathrm{est}}(l_{50})$ and $\bar{F}'_{\mathrm{est}}(l_{50})$ denote the RMS value and the linear mean value of $F'_{\mathrm{est}}(l_{50},z)$, respectively. $\hat{B}(l_{50})$ is smoothed with a moving median filter of length 71, resulting in $B(l_{50})$. Finally, the estimate of the time-dependent specific fluctuation strength $\hat{F}'(l_{50},z)$ is smoothed by using a lowpass filter of order one with a time constant of 0.75 s, resulting in the final estimate of the time-dependent specific fluctuation strength $F'(l_{50},z)$.

3.11 Calculation of time-dependent fluctuation strength

The time-dependent fluctuation strength $F(l_{50})$ is calculated by summing $F'(l_{50},z)$ over all 53 critical bands and taking the overlap $\Delta z = 0.5$ into account by multiplying the result with Δz. The representative fluctuation strength value F is calculated by taking the 90th percentile of the time-dependent fluctuation strength $F(l_{50})$, discarding the initial 36 values, i.e., for $0 \le l_{50} \le 35$.

4. CONCLUSIONS

The mathematical details of the improved fluctuation strength algorithm based on the Sottek Hearing Model that is proposed for the 4th edition of ECMA-418-2 are presented. The validation results for different kinds of synthetic and technical sounds are published in [6].

5. REFERENCES

[1] H. Fastl, “Fluctuation strength of modulated tones and broadband noise,” in *Hearing – physiological bases and psychophysics* ed. by R. Klinke, R. Hartmann, (pp. 282-288). Springer, Berlin, Heidelberg, 1983.

[2] H. Fastl, “Fluctuation strength of FM-tones,” in *Proc. of 11th ICA Paris*, vol. (3), pp. 123–126, 1983.

[3] H. Fastl and E. Zwicker: *Psychoacoustics: Facts and models*, 3rd Edition, Springer, Berlin, Heidelberg, New York, 2006.

[4] R. Sottek, S. Anisovich, J. Becker and T. Lobato, “Perception of Fluctuating Sounds,” *in Proc. of DAGA 2021*, Wien, 2021.

[5] R. Sottek, T. Lobato and J. Becker, “Modeling the fluctuation strength of technical sounds,” *in Proc. of DAGA 2023*, Hamburg, 2023.

[6] R. Sottek, “Standardization of the Sottek Hearing Model-based Fluctuation Strength Calculation”, *in Proc. of Noise-Con 2025*, Stowe, 2025.

[7] ECMA-418: *Psychoacoustic metrics for ITT equipment – Part 2: methods for describing human perception based on the Sottek Hearing Model*, 3rd Edition, December 2024.

[8] R. Sottek: *Modelle zur Signalverarbeitung im menschlichen Gehör*, doctoral thesis, University of Aachen, 1993.

[9] R. Sottek and T. Lobato, “High-resolution spectral analysis (HSA) vs. discrete Fourier transform (DFT),” in *Proc. of INTER-NOISE and NOISE-CON 2021*, vol. (4), pp. 2555-2566, 2021.