

A Indicator Descriptions

Roughness (rough)

Range: [0,1]	Unit: Asper	Scale: linear
Signal repr.: mean binaural (FB)	Literature: [1, 2]	Implementation: The Mathworks Inc. Audio Toolbox [3]

Sharpness (sharp)

Range: [0,20]	Unit: Acum	Scale: linear
Signal repr.: mean binaural (FB)	Literature: [1]	Implementation: The Mathworks Inc. Audio Toolbox [3]

Fluctuation Strength (fluct)

Range: [0,1]	Unit: Vacil	Scale: linear
Signal repr.: mean binaural (FB)	Literature: [1]	Implementation: The Mathworks Inc. Audio Toolbox [3]

Periodic Modulation Frequency (modF)

Range: [0.1,100]	Unit: Hz	Scale: linear
Signal repr.: mean binaural (FB)	Literature:	Implementation: Own

The periodic amplitude modulation is calculated by means of a peak-finding algorithm that is applied to the spectrum X_{am} that consists of the Fourier transformation of the signal envelope (Hilbert transformation) as in

$$X_{am} = 20 \cdot \log_{10}(\mathcal{F}\{\mathcal{H}\{x\}\}). \quad (1)$$

The Fourier spectrum was logarithmically resampled with 96 taps per octave in a frequency range between 0.05 Hz and 20 Hz and the peak-finding algorithm was configured with minimum peak prominence of 6 dB, minimum peak height of $(\max\{X_{am}\} - 24 \text{ dB})$ and minimum peak distance of 1/3 octaves.

Periodic Modulation Depth (modDepthP)

Range: [-10,10]	Unit: None	Scale: linear
Signal repr.: mean binaural (FB)	Literature:	Implementation: Own

The modulation depth that is associated with the periodic modulation frequency is calculated as described in A and scaled such that a depth of 1 corresponds to equal modulation and signal amplitude.

Stochastic Modulation Depth (modDepthS)

Range: [-10,10]	Unit: None	Scale: linear
Signal repr.: mean binaural (FB)	Literature:	Implementation: Own

The stochastic modulation is calculated by subtracting up to three periodic modulation envelopes from the original time signal and the modulation depth represents the signal amplitude of this differenced signal.

Spectral Centroid (specCentroid)

Range: [80,12000]	Unit: Hz	Scale: log
Signal repr.: mean binaural (FB)	Literature: [4]	Implementation: The Mathworks Inc. Audio Toolbox [3]

Spectral Spread (specSpread)

Range: [0,1000]	Unit: Hz	Scale: linear
Signal repr.: mean binaural (FB)	Literature: [4]	Implementation: The Mathworks Inc. Audio Toolbox [3]

Spectral Skewness (specSkewness)

Range: [0,1]	Unit: None	Scale: linear
Signal repr.: mean binaural (FB)	Literature: [4]	Implementation: The Mathworks Inc. Audio Toolbox [3]

Spectral Kurtosis (specKurtosis)

Range: [0,10000]	Unit: None	Scale: linear
Signal repr.: mean binaural (FB)	Literature: [4]	Implementation: The Mathworks Inc. Audio Toolbox [3]

Spectral Entropy (specEntropy)

Range: [0,1]	Unit: None	Scale: linear
Signal repr.: mean binaural (FB)	Literature: [5]	Implementation: The Mathworks Inc. Audio Toolbox [3]

Spectral Flatness (specFlatness)

Range: [0,1]	Unit: None	Scale: linear
Signal repr.: mean binaural (FB)	Literature: [6]	Implementation: The Mathworks Inc. Audio Toolbox [3]

Spectral Crest Factor (specCrest)

Range: [0.1,500]	Unit: None	Scale: log
Signal repr.: mean binaural (FB)	Literature: [4]	Implementation: The Mathworks Inc. Audio Toolbox [3]

Spectral Flux (specFlux)

Range: [$1e^{-12}$, $1e^{-1}$]	Unit: None	Scale: log
Signal repr.: mean binaural (FB)	Literature: [7]	Implementation: The Mathworks Inc. Audio Toolbox [3]

Spectral Slope (specSlope)

Range: [$-1e^{-6}$, $1e^6$]	Unit: None	Scale: lin
Signal repr.: mean binaural (FB)	Literature: [8]	Implementation: The Mathworks Inc. Audio Toolbox [3]

Spectral Derease (specDecr)

Range: [-10,1]	Unit: None	Scale: lin
Signal repr.: mean binaural (FB)	Literature: [4]	Implementation: The Mathworks Inc. Audio Toolbox [3]

Spectral Rolloff Point (specRolloff)

Range: [0.1,8000]	Unit: Hz	Scale: log
Signal repr.: mean binaural (FB)	Literature: [7]	Implementation: The Mathworks Inc. Audio Toolbox [3]

Octave Band Energy (oct)

Range: [-120,0]	Unit: dB	Scale: lin
Signal repr.: mean binaural (FB)	Literature: [9]	Implementation: The Mathworks Inc. Audio Toolbox [3]

Mel Frequency Cepstral Coefficient (mfcc)

Range: [-100,100]	Unit: None	Scale: linear
Signal repr.: mean binaural (FB)	Literature: [10]	Implementation: The Mathworks Inc. Audio Toolbox [3]

Timbral Booming (booming)

Range: [-20,20]	Unit: None	Scale: linear
Signal repr.: mean binaural (FB)	Literature: [11]	Implementation: AudioCommons [12]

Loudness (Zwicker) (loudZwick)

Range: [0.01,100]	Unit: Sone	Scale: log
Signal repr.: pressure (FB)	Literature: [2]	Implementation: The Mathworks Inc. Audio Toolbox [3]

Loudness (Moore-Glasberg) (loudMoore)

Range: [0.01,100]	Unit: Sone	Scale: log
Signal repr.: pressure (FB)	Literature: [13]	Implementation: The Mathworks Inc. Audio Toolbox [3]

Loudness units relative to full scale (momentary) (lufsMom)

Range: [-80,0]	Unit: LUFS	Scale: linear
Signal repr.: pressure (FB)	Literature: [14, 15]	Implementation: The Mathworks Inc. Audio Toolbox [3]

Loudness units relative to full scale (short-term) (lufsShort)

Range: [-80,0]	Unit: LUFS	Scale: linear
Signal repr.: pressure (FB)	Literature: [14, 15]	Implementation: The Mathworks Inc. Audio Toolbox [3]

Loudness units relative to full scale (integrated) (lufsInt)

Range: [-80,0]	Unit: LUFS	Scale: linear
Signal repr.: pressure (FB)	Literature: [14, 15]	Implementation: The Mathworks Inc. Audio Toolbox [3]

Loudness units relative to full scale (integrated) (lufsRange)

Range: [-10,10]	Unit: LU	Scale: linear
Signal repr.: pressure (FB)	Literature: [14, 15, 16]	Implementation: The Mathworks Inc. Audio Toolbox [3]

True Peak (lufsPeak)

Range: [-10,10]	Unit: LU	Scale: linear
Signal repr.: pressure (FB)	Literature: [15]	Implementation: The Mathworks Inc. Audio Toolbox [3]

L_A (la)

Range: [0,120]	Unit: dB	Scale: linear
Signal repr.: pressure (FB)	Literature: [17]	Implementation: The Mathworks Inc. Audio Toolbox [3]

$L_{A,eq}$ (laeq)

Range: [0,120]	Unit: dB	Scale: linear
Signal repr.: pressure (FB)	Literature: [17]	Implementation: The Mathworks Inc. Audio Toolbox [3]

$L_{A,peak}$ (lapeak)

Range: [0,120]	Unit: dB	Scale: linear
Signal repr.: pressure (FB)	Literature: [17]	Implementation: The Mathworks Inc. Audio Toolbox [3]

$L_{A,max}$ (lamax)

Range: [0,120]	Unit: dB	Scale: linear
Signal repr.: pressure (FB)	Literature: [17]	Implementation: The Mathworks Inc. Audio Toolbox [3]

Horizontal direction of arrival (doaAz)

Range: [-180,180]	Unit: degree	Scale: linear
Signal repr.: ambisonic (FB)	Literature: [18]	Implementation: Spherical Array Processing (Politis)[19]

Longitudinal direction of arrival (doaEl)

Range: [-90,90]	Unit: degree	Scale: linear
Signal repr.: ambisonic (FB)	Literature: [18]	Implementation: Spherical Array Processing (Politis)[19]

Diffuseness (diff)

Range: [0,1]	Unit:	Scale: linear
Signal repr.: ambisonic (FB)	Literature: [18]	Implementation: Spherical Array Processing (Politis)[19]

Interaural level difference (ild)

Range: [-30,30]	Unit: dB	Scale: linear
Signal repr.: binaural (FB)	Literature: [20]	Implementation: Auditory Modeling Toolbox (AMT) version 1.2[21]

Interaural time difference (itd)

Range: [-1,1]	Unit: ms	Scale: linear
Signal repr.: binaural (FB)	Literature: [20]	Implementation: Auditory Modeling Toolbox (AMT) version 1.2[21]

Interaural cross correlation (iacc)

Range: [0,1]	Unit:	Scale: linear
Signal repr.: binaural (FB)	Literature: [20]	Implementation: Auditory Modeling Toolbox (AMT) version 1.2[21]

Spherical Directivity Index (sphDi)

Range: [0,20]	Unit: dB	Scale: linear
Signal repr.: pressure (FB)	Literature: [22]	Implementation: Own

The calculation of the soundfield's directivity index follows Equation 2

$$DI = 10 \cdot \log_{10} \left(\frac{|p_{ref}|^2}{\int_{\varphi=0}^{2\pi} \int_{\theta=0}^{\pi} |p(\varphi, \theta)|^2 \sin \theta d\theta d\varphi} \right), \quad (2)$$

where $p_{ref} = \max(|p(\varphi, \theta)|^2)$. The directional sound pressure $p(\varphi, \theta)$ is calculated as plane wave decomposition of the incoming sound with 1° angular resolution.

Vertical directivity Index (sphDiEl)

Range: [0,20]	Unit: dB	Scale: linear
Signal repr.: pressure (FB)	Literature: [22]	Implementation: Own

The calculation of the soundfield's vertical directivity index follows Equation 3

$$DI_v = 10 \cdot \log_{10} \left(\frac{|p_{v,ref}|^2}{\int_{\theta=0}^{\pi} |p_v(\theta)|^2 \sin \theta d\theta} \right), \quad (3)$$

where $p_{v,ref} = \max(|p_v(\theta)|^2)$ and $p_v(\theta) = \int_{\varphi=0}^{2\pi} p(\varphi, \theta) d\varphi$. The directional sound pressure $p(\varphi, \theta)$ is calculated as plane wave decomposition of the incoming sound with 1° angular resolution.

Horizontal directivity Index (sphDiAz)

Range: [0,20]	Unit: dB	Scale: linear
Signal repr.: pressure (FB)	Literature: [22]	Implementation: Own

The calculation of the soundfield's horizontal directivity index follows Equation 4

$$DI_h = 10 \cdot \log_{10} \left(\frac{|p_{h,ref}|^2}{\int_{\varphi=0}^{2\pi} |p_h(\varphi)|^2 d\varphi} \right), \quad (4)$$

where $p_{h,ref} = \max(|p_h(\varphi)|^2)$ and $p_h(\varphi) = \int_{\theta=0}^{\pi} p(\varphi, \theta) \sin \theta d\theta$. The directional sound pressure $p(\varphi, \theta)$ is calculated as plane wave decomposition of the incoming sound with 1° angular resolution.

Spherical Pressure Ratio (sphPRatio)

Range: [-40,0]	Unit: dB	Scale: linear
Signal repr.: ambisonic (FB)	Literature:	Implementation: Own

The spherical pressure ratio describes the acoustic energy distribution between the Ambisonics signals for order $n = 0$ and $n > 0$. It is calculated as shown in Equation 5

$$s_p = 20 \cdot \log_{10} \left(\frac{|x_0|^2}{\frac{1}{(N+1)^2} \sum_{n=1}^N |x_n|^2} \right) \quad (5)$$

where N denotes the maximum order of the Ambisonics signal representation.

Spherical Gradient Ratio (sphGRatio)

Range: [-40,0]	Unit: dB	Scale: linear
Signal repr.: ambisonic (FB)	Literature:	Implementation: Own

The spherical pressure ratio describes the acoustic energy distribution between the Ambisonics signals for order $n = 0$ and $n > 0$. It is calculated as shown in Equation 6

$$s_g = 20 \cdot \log_{10} \left(\frac{\frac{1}{(N+1)^2-1} \sum_{n=1}^N |x_n|^2}{\frac{1}{(N+1)^2} \sum_{n=0}^N |x_n|^2} \right) \quad (6)$$

where N denotes the maximum order of the Ambisonics signal representation.

Spherical Gradient/Pressure Ratio (sphGPRatio)

Range: [-20,20]	Unit: dB	Scale: linear
Signal repr.: ambisonic (FB)	Literature:	Implementation: Own

The spherical pressure ratio describes the acoustic energy distribution between the Ambisonics signals for order $n = 0$ and $n > 0$. It is calculated as shown in Equation 7

$$s_{gp} = 20 \cdot \log_{10} \left(\frac{\frac{1}{(N+1)^2-1} \sum_{n=1}^N |x_n|^2}{|x_0|^2} \right) \quad (7)$$

where N denotes the maximum order of the Ambisonics signal representation.

References

- [1] Hugo Fastl and Eberhard Zwicker. *Psychoacoustics - Facts and models*. Springer Berlin, Heidelberg, 3rd editio edition, 2007. ISBN 3540231595. doi: 10.1007/978-3-540-68888-4.
- [2] ISO. ISO 532-1:2017 Acoustics - Methods for calculating loudness - Part 1: Zwicker method, 2017.
- [3] The Mathworks Inc. Audio Toolbox, 2022. URL <https://de.mathworks.com/products/audio.html>.
- [4] Geoffroy Peeters. A large set of audio features for sound description (similarity and classification) in the CUIDADO project. white paper, 2004. URL http://recherche.ircam.fr/anasyn/peeters/ARTICLES/Peeters_2003_cuidadoaudiofeatures.pdf.
- [5] Hemant Misra, Shajith Ikbal, Hervé Boursard, and Hynek Hermansky. Spectral entropy based feature for robust ASR. In *IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP*, 2004. doi: 10.1109/icassp.2004.1325955.
- [6] Aggelos Pikrakis, Theodoras Giannakopoulos, and Sergios Theodoridis. A Speech/Music Discriminator of Radio Recordings Based on Dynamic Programming and Bayesian Networks. *IEEE Transactions on Multimedia*, 10(5):846–857, 2008. doi: 10.1109/TMM.2008.922870.
- [7] Eric Scheirer and Malcolm Slaney. Construction and Evaluation of a Robust Multifeature Speech/Music Discriminator. *IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP*, 2:1331–1334, 1997. doi: 10.1109/icassp.1997.596192.
- [8] Alexander Lerch. *An Introduction to Audio Content Analysis*. John Wiley & Sons, Inc., Hoboken, NJ, 2012. ISBN 9781118393550. doi: 10.1002/9781118393550.
- [9] Acoustical Society of America and American National Standards Institute. ANSI S1.11: Specification for Octave-Band and Fractional-Octave-Band Analog and Digital Filters. *American National Standards on Acoustics*, 1986(734), 1986.
- [10] Steven B. Davis and Paul Mermelstein. Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences. *Transactions on Acoustics, Speech, and Signal Processing*, 28(4):357–366, 1980.
- [11] Shigeko Hatano and Takeo Hashimoto. Booming index as a measure for evaluating booming sensation. *Proc. of Internoise*, pages 4332–4336, 2000.
- [12] AudioCommons and Institute of Sound Recording. Timbral Models, 2019. URL <https://www.audiocommons.org/>.
- [13] ISO. ISO 532-2:2017 Acoustics - Methods for calculating loudness - Part 2: Moore-Glasberg method, 2017.
- [14] EBU. EBU - R 128: Loudness normalisation and permitted maximum level of audio signals, 2014.
- [15] ITU. ITU-R BS.1770-4: Algorithms to measure audio programme loudness and true-peak audio level, 2015.
- [16] EBU. EBU Tech 3342: Loudness Range: A Measure to Supplement EBU R 128 Loudness Normalization, 2016.
- [17] IEC. IEC 61672-1:2013 Electroacoustics - Sound level meters - Part 1: Specifications, 2013.
- [18] ville Pulkki. Spatial sound reproduction with directional audio coding. *J. Audio Eng. Soc.*, 55(6):503–516, 2007.

- [19] Archontis Politis. *Microphone array processing for parametric spatial audio techniques*. PhD thesis, Aalto University, Finland, 2016. URL <https://github.com/polarch>.
- [20] Jens Blauert. *Spatial Hearing*. The MIT Press, 1997. ISBN 0-262-02413-6.
- [21] Piotr Majdak, Clara Hollomey, and Robert Baumgartner. AMT 1.x: A toolbox for reproducible research in auditory modeling. *Acta Acustica*, 6(19), may 2022. ISSN 2681-4617. doi: 10.1051/aacus/2022011. URL <https://www.amtoolbox.org/>.
- [22] IEC. IEC 60268-5:2003 Sound system equipment – Part 5:Loudspeakers, 2003.