OPEN ACCESS Freely available online

J. Bergner, D. Schössow, S. Preihs and J. Peissig, "Identification of Discriminative Acoustic Dimensions in Stereo, Surround and 3D Music Reproduction"
*J. Audio Eng. Soc.*, vol. 71, no. 7/8, pp. 420–430, (2023 July/August).
DOI: https://doi.org/10.17743/jaes.2022.0071



# Identification of Discriminative Acoustic Dimensions in Stereo, Surround and 3D Music Reproduction

**JAKOB BERGNER,* DAPHNE SCHÖSSOW,**
(jakob.bergner@ikt.uni-hannover.de) (daphne.schoessow@ikt.uni-hannover.de)

**STEPHAN PREIHS, AND JÜRGEN PEISSIG**
(stephan.preihs@ikt.uni-hannover.de) (peissig@ikt.uni-hannover.de)

*Institute of Communications Technology, Leibniz University Hannover, Germany*

This work is motivated by the question of whether different loudspeaker-based multichannel playback methods can be robustly characterized by measurable acoustic properties. For that, underlying acoustic dimensions were identified that allow for a discriminative sound field analysis within a music reproduction scenario. The subject of investigation is a set of different musical pieces available in different multichannel playback formats. Re-recordings of the stimuli at a listening position using a spherical microphone array enable a sound field analysis that includes, in total, 237 signal-based indicators in the categories of loudness, quality, spaciousness, and time. The indicators are fed to a factor and time series analysis to identify the most relevant acoustic dimensions that reflect and explain significant parts of the variance within the acoustical data. The results show that of the eight relevant dimensions, the dimensions "High-Frequency Diffusivity," "Elevational Diffusivity," and "Mid-Frequency Diffusivity" are capable of identifying statistically significant differences between the loudspeaker setups. The presented approach leads to plausible results that are in accordance with the expected differences between the loudspeaker configurations used. The findings may be used for a better understanding of the effects of different loudspeaker configurations on human perception and emotional response when listening to music.

## 0 INTRODUCTION

The use of multiple loudspeakers for the reproduction of music and other entertaining audio content has been established within various technologies over decades. One of the main motivations in developing technologies such as stereo, surround sound, or 3D audio is to enhance the spatial image and impression of the reproduced audio experience [1]. The evolution of these audio reproduction techniques does not only involve the number of loudspeakers in use but also their spatial arrangement. With the cinema industry as a technology driver, the higher technical expenditure in each case became affordable for home applications with only a slight delay. Starting from stereo with two loudspeakers, to quadrophonic sound with four loudspeakers to the commercially very successful 5.1 surround sound with five loudspeakers and an additional subwoofer up to 7.1 surround sound with seven loudspeakers plus subwoofer, the number of loudspeakers increases analogously to the industry's promises of spatial imaging within the listening plane. With the incorporation of additional elevated loudspeakers, terms such as 3D audio or immersive audio become more and more widespread in technology and marketing, that comprise both audio rendering algorithms as well as loudspeaker setups, such as 22.2 surround sound. The additional height layer of loudspeakers is promised to further increase listener envelopment and spatial plausibility. There exists a mentionable body on research of perceptual effects provoked by these technologies, e.g., in [2–4]. However, at the same time, little information is available what properties of the reproduced sound field actually change with different reproduction technologies and if the perceptual effects can be explained or modelled with acoustic terms. Thus, this work proposes a methodology to identify statistically relevant underlying acoustic dimensions of music reproduction with different loudspeaker setups. These acoustic dimen-

*To whom correspondence should be addressed, e-mail: (jakob.bergner@ikt.uni-hannover.de). Last updated: January 13, 2023

Table 1. Positions of the loudspeakers used in spherical coordinates (origin at head position) with azimuth φ [−180°, 180°], elevation θ [−90°, 90°], and distance *d* in meters.

| | L | R | C | LS | RS | TL | TR | TLS | TRS |
|---|---|---|---|---|---|---|---|---|---|
| φ | 30 | -30 | 0 | 145 | −145 | 30 | -30 | 146 | −146 |
| θ | 0 | 0 | 0 | 0 | 0 | 23 | 22 | 25.5 | 24.5 |
| *d* | 2.50 | 2.50 | 2.20 | 2.20 | 2.20 | 2.94 | 2.98 | 2.75 | 2.72 |

sions are then evaluated for four exemplary loudspeaker layouts, namely mono, stereo, 5.1 surround sound, and 5.1.4 surround sound. Each layout studied adds spatial direction to the speaker positioning, so this selection of formats can serve as an example for other and/or more advanced speaker configurations. Pieces of music that were explicitly produced for these loudspeaker setups were played back and the generated sound field is analyzed at the proposed listening position in terms of a large number of signal-based acoustic indicators. By means of multivariate methods, underlying latent dimensions are deduced and subsequently analyzed statistically. This procedure is known from other fields of application, e.g., for the development of fundamental perceptual dimensions in the evaluation of environments [5] where a large number of semantic items are condensed to few latent constructs. This approach was also applied in order to identify the two main affective qualities of soundscapes *valence* and *arousal* summarized in [6] that eventually found their way into standardization in ISO/TS 12913-1/2/3 [7]. In the field of audio signal processing, the aggregation of individual indicators into abstract characterizing constructs is also known as "Audio Fingerprinting" [8], which is used to index audio files for automatic search algorithms in commercially available applications. The presented work is a detailed continuation and refinement of [9].

## 1 TECHNICAL INFRASTRUCTURE

For the acoustic analysis, the music reproduction was performed with a loudspeaker system within a listening room of approximately 30 $m^2$ that is acoustically treated for optimized audio reproduction with reverberation time between 0.2 and 0.3 s in compliance with the suggestions from ITU-R BS.1116-3 [10] regarding background noise, reflection pattern, and loudspeaker equalization. A detailed description of the acoustic properties of the listening room can be found in [11]. The loudspeaker system consists of nine full-range speakers Neumann KH 120 A and two subwoofers Neumann KH 810 G with a crossover frequency of 60 Hz. The positioning of the speakers is in agreement with ITU-R BS.2051-2 [12] as shown in Table 1.

The loudspeaker system was equalized toward equal gain and delay (distance compensation) and minimum frequency response deviation. The target frequency response is motivated by ITU-R BS.1116-3 [10]; however, the recommendation's low-frequency roll-off was neglected, and instead, a more natural room gain concept for low frequencies was followed [13]. This idea aims to preserve an expected low-frequency behavior for loudspeaker reproduction in rooms due to less low-frequency absorption compared with high-frequency energy. The room gain was estimated by modeling the low-frequency portion up to 200 Hz of the average room transfer functions from 42 loudspeakers to 5 microphone positions by means of a shelving filter. Equalization was applied by means of 8 biquad parametric peak/notch filters aiming to compensate an average of room responses at three microphone positions (center and 8.5 cm to the left and right) within a small listening area of the potential head location comparable to the procedure described in [14]. Furthermore, the octave around the crossover frequency between 42 Hz and 84 Hz was optimized with an allpass for each main loudspeaker so that the resulting frequency response deviates minimally from the target response. All equalization parameters were obtained by a global optimization procedure with constraints such that the phase response does not interfere with the reproduction in a negative way. The resulting room responses in third-octave bands can be taken from Fig. 1 for the ear-level loudspeakers (a) and top-level speakers (b).

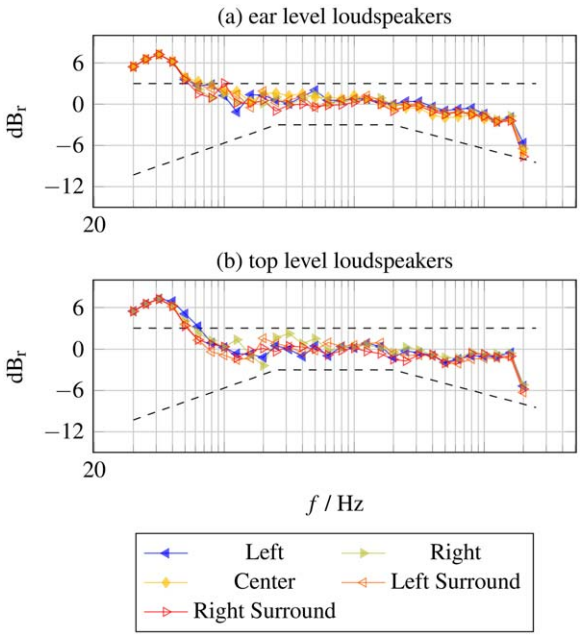


Fig. 1. Relative room responses in third-octave bands for ear level loudspeakers (a) and top level speakers (b).

## 2 STIMULI

The stimulus set comprises 8 excerpts of musical pieces of varying genre, ensemble size, and recording/production technique. Each piece of music is available in four versions of different channel-based loudspeaker reproduction formats: mono (center loudspeaker), stereo (left + right loudspeakers), 2D (5.1 surround sound), and 3D (5.1.4 surround sound). For the production of the stimuli, two audio engineers with experience in multichannel mixing were engaged to produce three equally sounding mixes (stereo, 2D, and 3D) from provided multitrack recordings that only vary in their spatial distribution but not in their general aesthetic characteristics. In practice, this was accomplished by man-

Table 2. Overview of investigated musical pieces.

| Label | Piece | Duration [s] | Genre and Orchestration | Production |
|---|---|---|---|---|
| Laudate | Laudate Dominum (Josep Vila) | 33.4 | A cappella choir: 12 singers (soprano, alto, tenor, bass) | 3D microphone setup + support microphones |
| Mellow | In a Mellow Tone (Janna Berger) | 35.4 | Jazz band: drum set, double bass, piano, female voice | 3D microphone setup + support microphones |
| Wunderschoen | Im wunderschönen Monat Mai (Robert Schumann) | 38.6 | Classic song: male voice, pf | 3D microphone setup + support microphones |
| School | School's Out (live; Alice Cooper) | 57.5 | Full live rock band | spot microphones + 3D ambience |
| Bilder | Pictures of an exhibition (Mussorgsky) | 37.3 | Large Orchestra | 3D microphone setup + support microphones |
| Walkuere | Ride of the Valkyries (Wagner) | 62.5 | Opera: Orchestra, female voices | manual upmix from commercial 5.1 content |
| Hantel | Die Hantel (Zweitaktmotor) | 61.8 | Electropop: synthesizers, male and female voices | pure multitrack studio production |
| Rokoko | Rokoko Variations (Tchaikovsky) | 68.1 | Classic chamber music: cello, woodwind quintet | manual upmix from commercial 5.1 content |

ual downmixes (or upmixes in two cases) of a common aesthetic 2D or 3D mix. The respective mono version was deduced from the stereo version by averaging left and right channel. An overview of the stimuli can be found in Table 2.

The loudness of the stimuli within the four playback formats was calibrated to minimize the median deviation of the short-term loudness units relative to full scale (LUFS) (EBU R 128 [15]) time series from the stereo reference. This procedure was validated by means of acoustic loudness measures. Monophonic sound pressure levels $L_{Aeq}$ and loudness according to ISO 532-1 [16] were measured with a Beyerdynamic MM1 microphone at the center of the listening area and binaural $L_{Aeq}$ as well as loudness according to ISO 532-2 [17], respectively, with a G.R.A.S 45BC-12 KEMAR. The comparison of loudness and level distributions between the formats revealed minor differences in dependence of the respective musical piece; however, no systematic and unexpected differences could be found. Another stage of validation was performed perceptually by an experienced audio engineer. In order to use the calibrated stimulus set for future listening tests, the overall loudness between the individual pieces of music was adjusted subjectively by the same audio engineer, aiming for plausibility in the reproduction of music with different ensemble sizes and genres.

## 3 IDENTIFICATION OF ACOUSTIC DIMENSIONS

The proposed methodology describes a process to obtain fundamental acoustic dimensions that are suitable to characterize and compare acoustic environments. The approach to identify those dimensions is data-driven and based on multivariate statistical analysis: A large number of observations of a large number of indicators are fed into appropriate methods for dimensionality reduction and variance maximization. The aim is to deduce dimensions that are interpretable with common acoustic terminology. The approach is exploratory, i.e., no hypothesis is formulated initially on how the dimensions must be composed of to be capable of discriminating acoustic environments generated with different loudspeaker setups. However, it can be assumed that this acoustic characteristic is reflected in terms of statistical variance within the acoustic data.

### 3.1 Acoustic Indicators

The indicators on which the dimension development is based is a collection of well-known and established signal parameters gathered from the fields of soundscape studies, music information retrieval, psychoacoustics, sound field analysis, and noise assessment. In fact, all of the indicators are documented to be used in modeling attempts of certain characteristics of general acoustic environments, such as soundscape quality, annoyance, and computer-aided classification and detection of specific sound event and environment classes as elaborated in [18] and its supplementary material. They are assigned to the three a priori categories loudness, quality, and spaciousness as listed in the following, with information on the respective conceptual reference and implementation in parentheses:

- **Quality:** Mel-frequency cepstral coefficients (MFCC) (reference: [19]; implementation: [20]), Spectral Centroid ([21];[20]), ∼ Crest ([21];[20]) Factor, ∼ Decrease ([21];[20]), ∼ Entropy ([22];[20]), ∼ Flatness ([23];[20]), ∼ Flux ([24];[20]), ∼ Kurtosis ([21];[20]), ∼ Roll-Off ([24];[20]), ∼ Skewness ([21];[20]), ∼ Spread ([21];[20]), Timbral Booming ([25];[26]), Roughness, Sharpness, Fluctuation Strength (all [27, 16];[20])
- **Loudness:** Sound pressure level (SPL) (A-/Z-weighting), octave band energy (all [28];[20]), loudness (Zwicker [16], Moore-Glasberg [17]; [20]), LUFS ([15, 29];[20])
- **Spaciousness:** Inter-aural level differences (ILD) ([30];[31]), inter-aural time differences (ITD), inter-aural cross correlation (IACC) (both [32];[31]), direction of arrival (horizontal, vertical), diffuseness (all [33];[34]), directivity index (regarding azimuth, elevation and full sphere) ([35];[18])

The indicators of the categories loudness and spaciousness were calculated both for a broadband frequency range

as well as for 10 individual octave bands with center frequencies ranging from 31 Hz to 16 kHz. The indicators themselves are calculated on the basis of one of three signal representations of the stimuli. Loudness and quality indicators are either calculated from a monophonic pressure representation or binaural representation, while the spaciousness indicators require binaural and Ambisonics signals. In order to analyze the resulting sound field present at the listening location, a re-recording of all stimuli by means of an mh acoustics em32 Eigenmike was performed. From the encoded Ambisonics signal representation generated by the mh acoustics EigenUnit-em32-Encoder, a binaural version was deduced by means of the BinauralDecoder of the IEM plugin suite [36] (fourth order), which implements a dual-band magnitude least-square binaural rendering approach for Ambisonics signals [37] as convolution with preprocessed head-related transfer functions measured with the Neumann KU-100 dummy head [38]. The direction of arrival and diffuseness indicators are calculated from the first-order Ambisonics representation as suggested in the DirAC approach [33], and the directivity indices are calculated as fourth-order plane wave decomposition with 1° angular resolution. The third monophonic sound pressure representation is finally generated from the 0th-order Ambisonic signals [33]. It has to mentioned that the two highest frequency bands with center frequencies $f_c = 8$ kHz and $f_c = 16$ kHz are potentially subject to spatial aliasing caused by the discrete sampling of the individual microphone capsules and the radius they are placed on [39, 40]. The used Ambisonics encoder aims to reduce this effect with help of a designated high-frequency extension [39] that helps to maintain an undistorted frequency response but that alters the spatial information in a nondetermined way at the same time. Thus, we can assume proper spectral and energetic characteristics within these frequency bands but unreliable spatial information (regarding both the binaural and the 3D sound field representation). However, because all stimuli are recorded, encoded, and processed in the same way, these artifacts should occur systematically and not between the stimuli in the subsequent analysis, which is why these frequency bands are kept for further investigations. In general, the parametrization of all indicator calculations (e.g., encoding and decoding of Ambisonics and binaural signals, choice of frequency bands and time-frequency resolution) may lead to slightly differing indicator values.

All indicators were calculated as time series with window length $l_w = 0.1$ s and hop size of $l_h = 0.05$ s. The time series were scaled to an expected value range, and logarithmic sampling was applied where necessary. Finally, a $z$-standardization (zero mean; unit variance) was applied to all indicators individually. In total, the input data consists of 237 indicators and 28,696 observations each.

### 3.2 Multivariate Statistics

The indicator's time series were then subject to multivariate analysis methods, specifically to factor analysis (FA) [41]. FA assumes that underlying latent factors become manifest in observed indicators, as shown in Fig. 2.

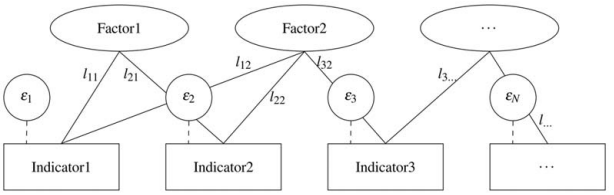


Fig. 2. Concept of FA with loadings $l_{ij}$ and unique variances $\varepsilon_i$.

FA can be used to transform data from the original space into an optimized space of latent dimensions. In mathematical terms, FA can thus be described in a generative way as Eq. (1)

$$\boldsymbol{X} = \boldsymbol{L} \cdot \boldsymbol{Y} + \epsilon, \tag{1}$$

where $\mathbf{X}$ is a $[N_{\text{ind}} \times N_{\text{obs}}]$ matrix of the original data, $\boldsymbol{L}$ is a specific loading matrix of dimension $[N_{\text{fac}} \times N_{\text{ind}}]$, $\boldsymbol{Y}$ is the resulting factor scores of $[N_{\text{fac}} \times N_{\text{obs}}]$, and $\varepsilon$ is a diagonal matrix of unique variances where $N_{\text{obs}}$ represents the number of observations, $N_{\text{ind}}$ the number of indicators, and $N_{\text{fac}}$ the number of factors. This equation is solved by means of an iterative estimation that minimizes the unique variance $\varepsilon$ using maximum likelihood and takes covariance $\text{Cov}(\boldsymbol{X})$ considerations into account, as formulated in [42] and implemented in [43]. The loading matrix comprises the individual weights of each indicator into each factor. The sum over rows, i.e., among indicators, yields the sum of squared loadings or the explained variance of a certain factor

$$s_j^2 = \sum_{i=1}^{N_{\text{ind}}} l_{ji}^2. \tag{2}$$

The relative loading represents the direction of the transformation and can be described as

$$\boldsymbol{L_{rel}} = \boldsymbol{L} \cdot \text{diag}\{\mathbf{s}\}^{-1}. \tag{3}$$

In FA, it is an important decision how many factors to keep, i.e., in this case, underlying acoustic dimensions. The Kaiser criterion assumes factors with $s_j^2 \geq 1$ to be relevant because they inhibit more variance than a single indicator. However, the parallel analysis according to Horn [44] is a more convenient method because it compares the explained variance with a random sample of the same size by means of a Monte Carlo simulation. The results for both criteria can be found in Table 3 for pure and *varimax* rotated FA.

In this work, we follow the parallel analysis suggestion for *varimax* rotated FA and keep the eight most prominent factors. Their explained variance portion can be taken as scree plot from Fig. 3 (top), and the associated loading matrix $\boldsymbol{L}$ is visualized in Fig. 3 (bottom).

Table 3. Relevant number of factors $N_{rel}$ according to Kaiser's criterion and parallel analysis for pure and rotated FA.

| | Kaiser Criterion | | Parallel Analysis | |
|---|---|---|---|---|
| | $N_{rel}$ | cumulative $s_j^2$ | $N_{rel}$ | cumulative $s_j^2$ |
| FA | 20 | 169.27 (71.42%) | 7 | 141.02 (59.50%) |
| FA *varimax* | 26 | 173.11 (73.04%) | 8 | 139.35 (58.80%) |

Table 4. Indicator composition of the first eight relevant factors $j$ with respective explained variance $s_j^2$ and relative loadings $l_{\text{rel},\,ij}$ in parentheses. Trailing numbers of the indicators denote the frequency bands. $N_{\text{ind},\,j}$ denotes how many indicators account for ≥51% of the factor's explained variance.

| Factor$j$ | $s_j^2$ | $N_{\text{ind},\,j}$ | Indicators |
|---|---|---|---|
| 1 | 69.74 (29.4%) | 40 | LAeq(0.116), LA(0.116), loudnessZwickerBands05(0.116), LApeak(0.116), LAeqBands06(0.115), loudnessZwicker(0.115), LABands06(0.115), LAeqBands05(0.115), LAmax(0.115), LApeakBands05(0.114), loudnessZwickerBands06(0.114), lufsMomBands06(0.114), LABands05(0.114), LAmaxBands06(0.113), lufsPeakBands05(0.113), lufsPeakBands06(0.113), LAeqBands07(0.113), LApeakBands06(0.113), LABands07(0.113), LAmaxBands05(0.113), lufsMomBands05(0.113), loudnessZwickerBands04(0.112), oct07(0.112), mfcc00(0.112), lufsMomBands07(0.111), oct06(0.111), LApeakBands04(0.111), LAmaxBands07(0.110), oct08(0.110), LAeqBands04(0.110), lufsMom(0.110), LABands04(0.109), lufsPeakBands04(0.109), lufsPeakBands07(0.108), lufsPeak(0.108), LAmaxBands04(0.107), lufsMomBands04(0.107), loudnessZwickerBands07(0.107), LApeakBands07(0.105), oct05(0.104) |
| 2 | 30.81 (13.0%) | 26 | spectralCentroid(−0.163), spectralDecrease(0.161), lufsMomBands00(0.150), lufsShortBands00(0.149), lufsMomBands01(0.147), lufsPeakBands01(0.147), oct02(0.147), lufsShortBands01(0.146), LAmaxBands00(0.146), lufsPeakBands00(0.145), LABands00(0.143), spectralRolloffPoint(−0.143), oct01(0.142), LAeqBands00(0.142), LAmaxBands01(0.142), LABands01(0.139), LApeakBands00(0.139), LAeqBands01(0.137), LApeakBands01(0.136), lufsShortBands02(0.135), oct00(0.130), lufsMomBands02(0.129), oct03(0.124), lufsPeakBands02(0.123), booming0(0.122), loudnessZwickerBands01(0.118) |
| 4 | 13.63 (5.8%) | 9 | sphDIAz08(−0.250), sphDIAz07(−0.248), diff08(0.246), diff07(0.245), sphDIAz09(−0.243), sphDI08(−0.239), sphDI07(−0.235), sphDI09(−0.234), sphDIAz06(−0.232) |
| 5 | 6.93 (2.9%) | 5 | fluct06(0.330), fluct05(0.329), fluct04(0.326), fluct07(0.319), fluct03(0.277) |
| 7 | 5.76 (2.4%) | 5 | rough05(0.349), rough06(0.347), rough07(0.341), rough04(0.335), rough08(0.319) |
| 3 | 4.80 (2.0%) | 21 | mfcc01(−0.306), sharp(0.231), booming2(−0.172), lufsShortBands09(0.153), lufsShortBands08(0.151), lufsMomBands09(0.147), lufsPeakBands09(0.146), oct04(−0.143), lufsPeakBands08(0.141), lufsMomBands08(0.140), LABands08(0.131), LABands09(0.131), lufsPeakBands03(−0.130), LABands03(−0.130), oct05(−0.128), loudnessZwickerBands03(−0.127), booming0(−0.127), LAeqBands03(−0.125), lufsMomBands03(−0.124), LApeakBands08(0.120), LAmaxBands08(0.120) |
| 8 | 4.15 (1.7%) | 5 | sphDIEl08(−0.344), sphDIEl09(−0.341), doaEl07(0.320), doaEl08(0.319), sphDIEl07(−0.310) |
| 10 | 3.52 (1.5%) | 3 | sphDI04(−0.449), sphDIEl04(−0.431), sphDIAz04(−0.351) |

In order to interpret the resulting factor scores, the composition of the individual factors is analyzed. Table 4 lists those indicators that load gravely to the respective factors. Only those indicators are retained whose sum of squared relative loadings accounts for ≥51% of a factor's explained variance and which, thus, have a characterizing effect. The number of indicators necessary to achieve this is shown as $N_{\text{ind},\,j}$.

The composition of the factors allow for a good interpretation and the possibility to describe them semantically. Factor 1 comprises loudness and level indicators for broadband and mid–high frequency range (frequency bands 4–7, respectively $f_c = 500\ldots4{,}000$ Hz). Factor 2 is dominated by spectral metrics as well as low-frequency loudness and levels (frequency bands 0–2, resp. $f_c = 31\ldots125$ Hz). The next factor 4 is composed of spherical directivity indices over the full sphere ("sphDI") and regarding the azimuthal plane ("sphDIAz"), as well as the DirAC diffuseness ("diff") based on the 3D intensity vector [33]. These indicators are apparent for frequency bands 6–9 (resp. $f_c = 2\ldots16$ kHz) and represent the high-frequency range. As discussed here, the frequency bands 8 and 9 are potentially subject to spatial aliasing. The fact that the indicators of these bands correlate with their undistorted counterparts of bands 6 and 7 suggest that the potential aliasing errors do not degrade the common variance of these factors, which is an important finding for the following analysis. Factor 5 consists solely of fluctuation strength of the frequency bands 3–8 (resp. $f_c = 250\ldots5{,}000$ Hz), whereas factor 7 consists of roughness indicators for frequency bands with $f_c = 500\ldots8{,}000$ Hz. Factor 3 comprises 21 indicators representing high-frequency timbre (sharpness, level, and loudness for bands with $f_c = 8\ldots16$ kHz) as well as low-frequency characteristics (timbral booming, low-frequency MFCC, and loudness) and can be interpreted as ratio between high- and low-frequency aspects. Factor 8 includes the elevational directivity index ("sphDIEl") and direction of arrival regarding the elevation ("doaEl") and factor 10,

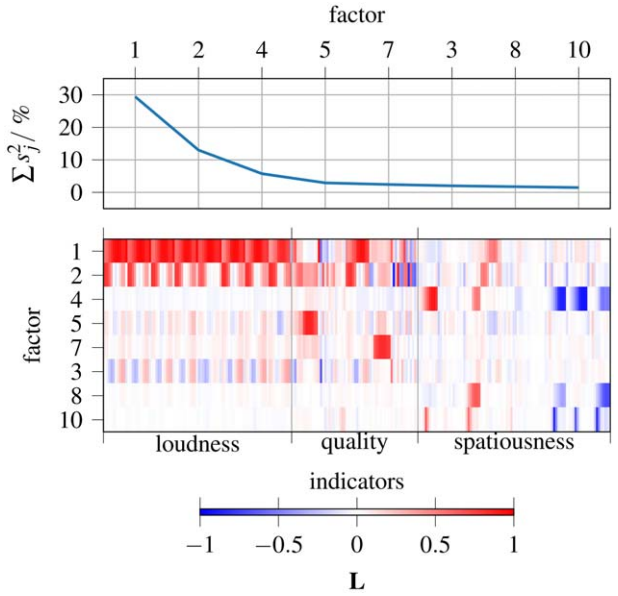


Fig. 3. Top: explained variance portion in % (scree plot), bottom: relative loading matrix $\boldsymbol{L}_{rel}$.

Table 5. Semantic descriptors of identified relevant factors forming underlying acoustic dimensions (cf. SEC. 3.2) as well as Friedman test statistics: p-values (statistically significant values are bold) and Kendall's coefficient of concordance W (cf. SEC. 4.1).

| Factor | Descriptor | $W$ | $p$ |
|---|---|---|---|
| 1 | "Loudness" | 0.28 | 0.08031 |
| 2 | "Low-Frequency Timbre" | 0.10 | 0.49364 |
| 4 | "High-Frequency Diffusivity" | **0.93** | **0.00006** |
| 5 | "Fluctuation" | 0.10 | 0.49364 |
| 7 | "Roughness" | 0.29 | 0.07032 |
| 3 | "High-Frequency Timbre" | 0.26 | 0.10454 |
| 8 | "Elevational Diffusivity" | **0.73** | **0.00055** |
| 10 | "Mid-Frequency Diffusivity" | **0.53** | **0.00559** |

the spherical directivity index regarding the full sphere, elevation, and azimuth ("sphDI", "sphDIEl", "sphDIAz") for the single-octave frequency band with $f_c = 500$ Hz. All in all, the eight most relevant factors can be described semantically as listed in Table 5. They finally form the underlying acoustic dimensions of music reproduction with different loudspeaker setups in this work.

## 4 RESULTS

### 4.1 Statistical Differences Within Acoustic Dimensions

The underlying acoustic dimensions are formed from the latent factors identified before. The expression of these dimensions is represented for each individual acoustic environment by the factor scores $\mathbf{Y}$ calculated with Eq. (1).

The distribution of the resulting factor scores can be found in Fig. 4. Each of the eight most relevant factors is shown individually, where factor scores (ordinate) for the individual loudspeaker setups (color coded) are grouped for each piece of music (abscissa). Outliers exist but are omitted in the visualization for clarity.

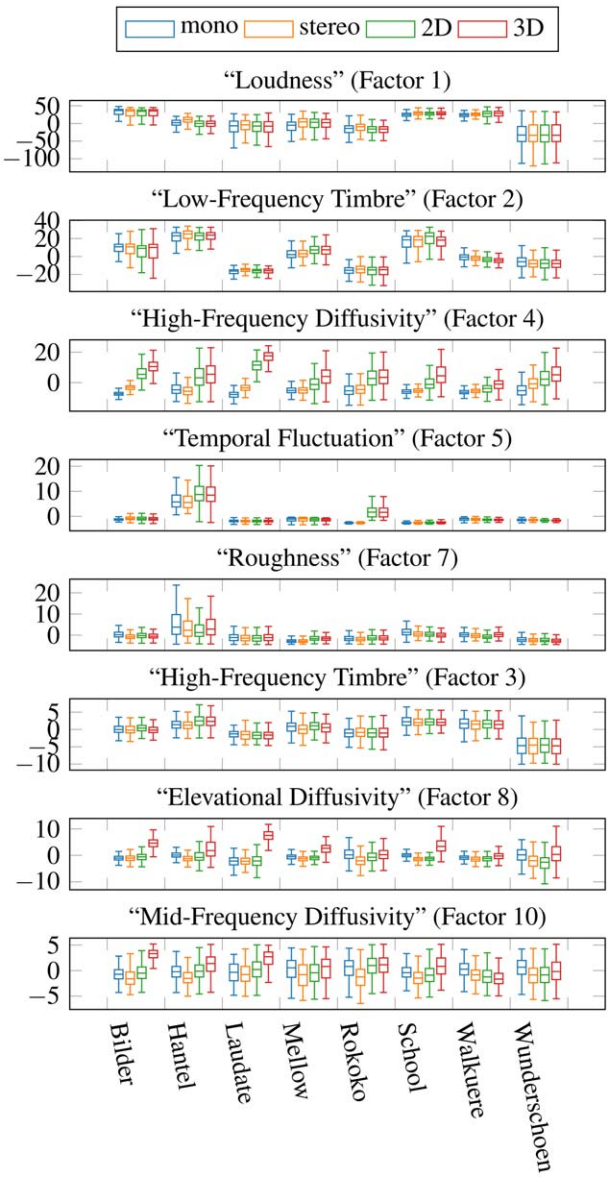


Fig. 4. Factor score distributions for the most relevant dimensions. "Loudness," "Low-Frequency Timbre," and "High-Frequency Timbre" show differences between pieces, whereas "High-Frequency Diffusivity" and "Elevational Diffusivity" show differences between loudspeaker setups.

From that, different patterns between the dimensions can be observed. For example, "Loudness," "Low-Frequency Timbre," and "High-Frequency Timbre" show differences between the musical pieces but seem to be stable between the loudspeaker setups. Other dimensions like "High-Frequency Diffusivity" and "Elevational Diffusivity" show distinct differences between loudspeaker setups but not that much between musical pieces.

In order to find underlying acoustic dimensions that actually change with the loudspeaker setup, appropriate statistics were applied. First, analysis of normality for each subgroup (distribution of factor scores for a specific musical piece and a specific loudspeaker setup) could not be asserted for all cases. For this purpose, Shapiro–Wilk tests were conducted and additionally validated with Q-Q plots to compensate their weakness for large sample sizes. Sub-

sequently, nonparametric methods were applied, namely, a Friedman test on ranks (as alternative for one-way repeated measure ANOVA), for testing the null hypothesis $H_0$: *"There is no difference in scores of a specific acoustic dimension between mono, stereo, 2D and 3D loudspeaker setups."* with a level of significance of $\alpha = 0.05$. The results of the Friedman test can be found in Table 5.

It shows highly significant factors 4, 8, and 10 ($p < 0.001$) with excellent, good, and moderate *Kendall's coefficient of concordance W*, respectively.

Post hoc paired Wilcoxon signed-rank tests (nonparametric alternative to paired *t*-tests) with one-sided alternative hypothesis $H_1$: *"The scores of a specific acoustic dimension and musical piece are greater for loudspeaker setup A compared to setup B."* were conducted for these three factors to examine differences between individual loudspeaker setups for each piece of music.

Table 6 shows the resulting *p*-values with Bonferroni adjustment for these tests. The columns denote the respective pairwise comparison, for example, whether the factor score distribution of the 3D condition is greater compared with the distribution of the 2D condition.

It can be seen that for factor 4 ("High-Frequency Diffusivity"), the condition with the higher number of loudspeakers exceeded the respective comparison condition in almost all cases. Obviously, an increase of the sound sources (loudspeakers) lead to higher factor scores within this dimension of diffusivity, which is an expected outcome.

A similar pattern can be found in factor 8 ("Elevational Diffusivity") especially for the comparisons of the 3D loudspeaker setups. This behavior also meets the expectations in such a way that additional elevated sound sources alter the directivity characteristic of the sound field regarding the elevation. However, the fact that the comparison between the 2D setup and the stereo setup also shows significant differences for six out of eight pieces of music in this dimension impairs the robustness of this dimension's interpretability. Furthermore, the distribution of this factor in Fig. 4 even shows a decreasing tendency for some pieces of music for mono, stereo, and 2D reproduction. This indicates that the elevational directivity index is sensitive to room acoustic influences such as floor and ceiling reflections, which might lead to higher factor scores for mono reproduction.

Factor 10 ("Mid-Frequency Diffusivity") exhibit 19 out of 48 significant pairwise comparisons, which is a somewhat ambiguous result. This ambiguity can also be seen in the distribution of this factor in Fig. 4 with large and overlapping interquartile ranges. A tendency that differences might be due to the higher number of loudspeakers can be observed; however, a content-dependent and nonsystematic relationship can not be denied. Thus, this factor must be interpreted as a vague measure for discriminating between the loudspeaker setups under test.

### 4.2 Temporal Characteristics

The presented methodology is based on similarities and differences of the distribution of factor scores. In principle, this would include the assumption that each short-term observation is independent of any other observation. This of course is not the case since time series are investigated that are tied to a process with both stochastic and deterministic features. Hence, in order to assure that the above made statements are valid not only for the distributions but also for the time series, a further analysis step was conducted. With the help of independent component analysis (ICA) [45] it is possible to detect underlying signal bases. The method assumes that observed signals are mixtures of superimposed basis signals. The decomposition of the four signal observations (mono, stereo, 2D, 3D) for each dimension and each piece of music into subcomponents is ought to reveal similarities in the temporal characteristic.

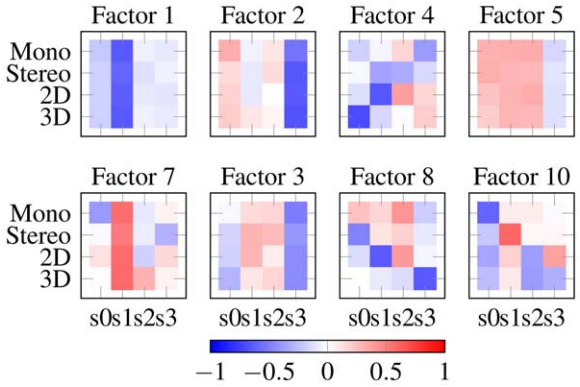


Fig. 5. Mixing matrix of the ICA for the piece Laudate denoting the composition of fundamental signal shapes $s0$, $s1$, $s2$, and $s3$ to represent the slopes of the factor score's time series of the respective loudspeaker setups.

Fig. 5 shows the exemplary mixing matrix of the ICA with four basis signal components $s0$, $s1$, $s2$, $s3$ of the piece Laudate. It can be seen that for the factors 1, 2, 5, 7, and 3, a single component is mixed with large weights to the time series of the factor scores of the respective loudspeaker setups. This vertical structure means that all four conditions are based on similar time series properties. The factors 4, 8, and 10 have different characteristics. Here such structures cannot be identified, which means that the time series of the four loudspeaker conditions do differ in a relevant way. These both findings of similarities and differences confirm the assumptions that not only the distributions but also the time series of the identified factor scores discriminate the four loudspeaker conditions within the factors 4, 8, and 10.

## 5 SUMMARY AND DISCUSSION

The proposed methodology to identify underlying acoustic dimensions of general acoustic environments was capable of revealing acoustic properties that discriminate different loudspeaker systems for music reproduction. The presented approach based on robust statistical treatment of signal parameters produced plausible and expected results but certain ambiguities at the same time: First, with proper calibration, the acoustic dimensions "Loudness," "Low-Frequency Timbre," "Fluctuation," "Rough-

Table 6. Paired Wilcoxon signed-rank tests with one-sided alternative (condition A is greater than condition B) for the factors found to be significantly different between loudspeaker setups. Significance is shown with bold typeset Bonferroni-adjusted $p$-values $p^* = p \cdot (8 \cdot 6)$ for $p^* < 0.05$ (*) and $p^* < 0.01$ (**), whereas non significance is denoted with n/s.

| Factor | Piece | 3D > 2D | 3D > Stereo | 3D > Mono | 2D > Stereo | 2D > Mono | Stereo > Mono |
|---|---|---|---|---|---|---|---|
| 4 | Bilder | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) |
| | Hantel | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | n/s (1.000) |
| | Laudate | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) |
| | Mellow | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | n/s (1.000) |
| | Rokoko | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) |
| | School | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) |
| | Walkuere | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) |
| | Wunderschoen | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) |
| 8 | Bilder | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | n/s (1.000) |
| | Hantel | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | n/s (1.000) | n/s (1.000) |
| | Laudate | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | n/s (1.000) | n/s (1.000) | n/s (1.000) |
| | Mellow | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | n/s (1.000) | n/s (1.000) |
| | Rokoko | ** (<**0.001**) | ** (<**0.001**) | n/s (1.000) | ** (<**0.001**) | n/s (1.000) | n/s (1.000) |
| | School | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | n/s (1.000) | n/s (1.000) |
| | Walkuere | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | n/s (1.000) | n/s (1.000) |
| | Wunderschoen | ** (<**0.001**) | ** (<**0.001**) | * (**0.040**) | n/s (1.000) | n/s (1.000) | n/s (1.000) |
| 10 | Bilder | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | * (**0.041**) | n/s (1.000) |
| | Hantel | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | ** (**0.002**) | n/s (1.000) |
| | Laudate | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | n/s (1.000) |
| | Mellow | ** (<**0.001**) | ** (<**0.001**) | n/s (0.315) | ** (<**0.001**) | n/s (1.000) | n/s (1.000) |
| | Rokoko | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | n/s (1.000) |
| | School | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | ** (<**0.001**) | n/s (1.000) | n/s (1.000) |
| | Walkuere | n/s (1.000) | n/s (1.000) | n/s (1.000) | n/s (1.000) | n/s (1.000) | n/s (1.000) |
| | Wunderschoen | ** (<**0.001**) | ** (<**0.001**) | n/s (1.000) | n/s (0.574) | n/s (1.000) | n/s (1.000) |

ness," and "High-Frequency Timbre" show no significant differences between the loudspeaker setups.

Second, the acoustic dimension "High-Frequency Diffusivity" (dimension 4) shows significant difference between loudspeaker setup according to an omnibus Friedman test. Furthermore, an increase of loudspeakers in use leads to an increase in scores of this factor, as shown by a post hoc Wilcoxon test.

Third, "Elevational Diffusivity" (dimension 8) is capable of discriminating the 3D setup from the others as shown in both Friedman and Wilcoxon test. At the same time, this factor delivers ambiguous results when comparing mono, stereo, and 2D where all loudspeakers are placed at the same height and room acoustic effects dominate.

Fourth, despite significant omnibus Friedman test results, "Mid-Frequency Diffusivity" (dimension 10) shows no systematic and robust post hoc test results for comparing the different loudspeaker setups. A tendency toward higher values with a larger number of loudspeakers used can be observed, but a content-dependent and/or arbitrary influence is highly present. This fact is also present in the low-variance explanation of 1.5 % and low Kendall's concordance $W = 0.53$.

In summary, it can be stated that diffusivity—a dimension indicating whether the incoming sound comes from a specific direction or diffusely from all sides—can be reliably calculated and can serve as a distinguishing feature for the investigated loudspeaker configurations. It may be questioned at this point if the findings can be generalized to either other pieces of music or other loudspeaker setups or even for general acoustic environments. Indication that the methodology of developing underlying acoustic dimensions itself is a suitable approach for general acoustic environments can be found in [18]. At the same time, it may be the case that a different set of acoustic stimuli produce different factor compositions during FA. This is because, strictly speaking, the characterizing loading matrix is only valid for the population of observations that was originally fed into it. In the present case and in similar cases, it is important to make a balanced selection of stimuli and observations with respect to the hypothesis being pursued.

After all, a potential application of the findings could be a contribution to the modeling of human perception. With the robust identification of different acoustic dimensions, it is now possible to investigate whether the effects of certain loudspeaker configurations on human auditory perception can be explained to some extent by these parameters. Appropriate listening tests with assessment of physiological, perceptual, and emotional responses were carried out and currently analyzed within the project Richard Wagner 3.0 [46]. If perceptual qualities such as envelopment, apparent source width, or even overall auditory experience are to be predicted using acoustic parameters, the dimensions of "High-Frequency Diffusivity," "Elevational Diffusivity," and "Mid-Frequency Diffusivity" form a set of attributes worth targeting. Apart from that, a general room acoustical characterization with or without electroacoustic system may be assessed in terms of the identified dimensions.

Finally, an exemplary comparison of all identified underlying acoustic dimensions can be found in Fig. 6. It shows

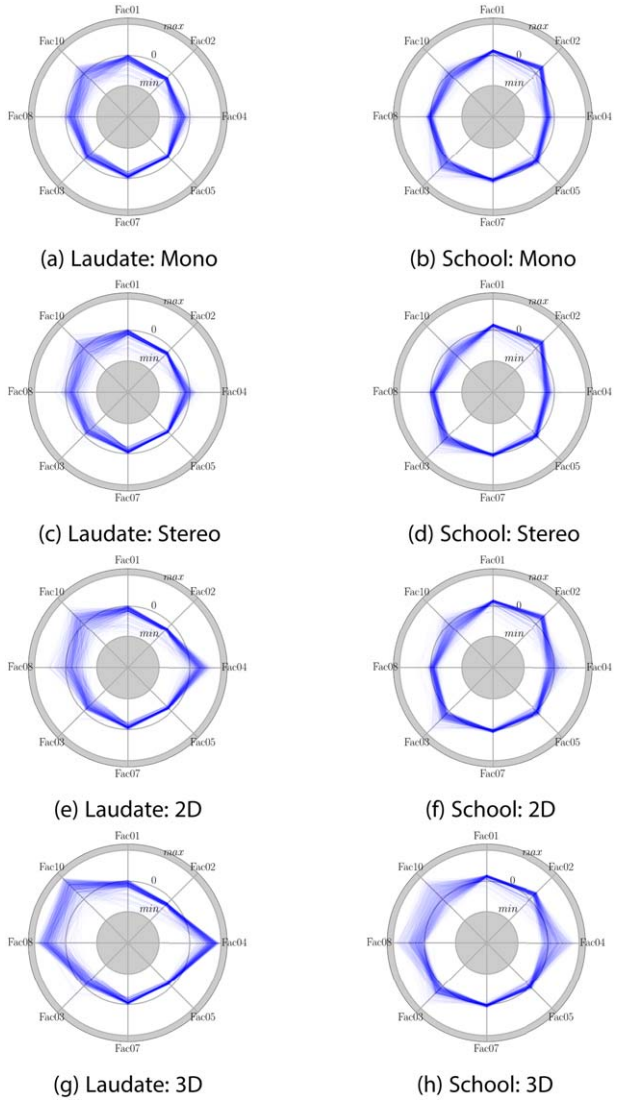


Fig. 6. Acoustic fingerprints of music pieces Laudate (left column) and School (right column) for the four loudspeaker setups. The acoustic dimensions are arranged clockwise starting from 12 o'clock: "Loudness," "Low-Frequency Timbre," "High-Frequency Diffusivity," "Fluctuation," "Roughness," "High-Frequency Timbre," "Elevational Diffusivity," and "Mid-Frequency Diffusivity."

an acoustic fingerprint of two exemplary pieces of music for all four loudspeaker setups. The polar axes of each fingerprint represent the respective factor or acoustic dimension. Each time window of 0.1 s is represented by a faint blue polar line, which also supports a general understanding of the temporal distribution of factor scores. This visualization allows us to intuitively compare general characteristics of the musical piece, as well as the progression of the dimensions between the loudspeaker setups.

## 6 ACKNOWLEDGEMENT

The authors are thankful for the research grant of the project Richard Wagner 3.0 funded by "Niedersächsisches Vorab," a joint program by the Volkswagen Foundation in conjunction with the Lower Saxony Ministry for Science and Culture (ZN3497). Further thanks goes to our project partners Yves Wycisk and Reinhard Kopiez from the Hanover Music Lab of the Hanover University of Music, Drama and Media for their efforts in compiling and revising the stimuli.

## 7 REFERENCES

[1] A. Roginska and P. Geluso (Eds.), *Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio* (Routledge, New York, 2018). https://doi.org/10.4324/9781315707525.

[2] C. Eaton and H. Lee, "Subjective Evaluations of Three-Dimensional, Surround and Stereo Loudspeaker Reproductions Using Classical Music Recordings," *Acoust. Sci. Technol.*, vol. 43, no. 2, pp. 149–161 (2022 Mar.). https://doi.org/10.1250/ast.43.149.

[3] M. Schoeffler, A. Silzle, and J. Herre, "Evaluation of Spatial/3D Audio: Basic Audio Quality Versus Quality of Experience," *IEEE J. Sel. Top. Signal Process.*, vol. 11, no. 1, pp. 75–88 (2017 Feb.). https://doi.org/10.1109/JSTSP.2016.2639325.

[4] E. Hahn, "Musical Emotions Evoked by 3D Audio," in *Proceedings of the AES International Conference on Spatial Reproduction - Aesthetics and Science* (2018 Jul.), paper P12-2.

[5] J. A. Russell, L. M. Ward, and G. Pratt, "Affective Quality Attributed to Environments - A Factor Analytic Study," *Environ. Behav.*, vol. 13, no. 3, pp. 259–288 (1981 May). https://doi.org/10.1177/0013916581133001.

[6] A. Fiebig, P. Jordan, and C. C. Moshona, "Assessments of Acoustic Environments by Emotions - The Application of Emotion Theory in Soundscape," *Front. Psychol.*, vol. 11 (2020 Nov.). https://doi.org/10.3389/fpsyg.2020.573041.

[7] ISO, "Acoustics — Soundscape — Part 2. Data Collection and Reporting Requirements," *DIN ISO 12913-2* (2018 Aug.).

[8] A. Lerch, *An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics* (John Wiley & Sons, Inc., Hoboken, NJ, USA, 2012). https://doi.org/10.1002/9781118393550.

[9] J. Bergner, D. Schössow, S. Preihs, Y. Wycisk, K. Sander, R. Kopiez, *et al.*, "Analyzing the Degree of Immersion of Music Reproduction by means of Acoustic Fingerprinting," in *Proceedings of the Fortschritte der Akustik - DAGA* (Stuttgart, Germany) (2022 Mar.).

[10] ITU-R, "Methods for the Subjective Assessment of Small Impairments in Audio Systems," *Recommendation ITU-R BS.1116-3* (2015 Feb.).

[11] R. Hupke, J. Ordner, J. Bergner, M. Nophut, S. Preihs, and J. Peissig, "Towards a Virtual Audiovisual Environment for Interactive 3D Audio Productions," in *Proceedings of the AES International Conference on Immersive and Interactive Audio* (2019 Mar.), paper 51.

[12] ITU-R, "Advanced Sound System for Programme Production," *Recommendation ITU-R BS.2051-2* (2018 Jul.).

[13] J. Abildgaard Pedersen and F. El-Azm,"Natural Timbre in Room Correction Systems (Part II)," in *Proceedings of the AES 32nd International Conference: DSP for Loudspeakers* (2007), paper 5.

[14] J. Bergner, S. Preihs, R. Hupke, and J. Peissig, "A System for Room Response Equalization of Listening Areas Using Parametric Peak Filters," in *Proceedings of the AES International Conference on Immersive and Interactive Audio* (2019 Mar.), paper 45.

[15] EBU-R, "Loudness Normalisation and Permitted Maximum level of Audio Signals," *Recommendation EBU-R 128* (2014 Jun.).

[16] ISO, "Acoustics — Methods for Calculating Loudness — Part 1: Zwicker Method," *Standard 532-1:2017* (2017 Jun.).

[17] ISO, "Acoustics — Methods for Calculating Loudness — Part 2: Moore-Glasberg Method," *Standard 532-2:2017* (2017 Jun.).

[18] J. Bergner and J. Peissig, "On the Identification and Assessment of Underlying Acoustic Dimensions of Soundscapes." *Acta Acust.*, vol. 6, paper 46 (2022 Oct.). https://doi.org/10.1051/aacus/2022042.

[19] S. B. Davis and P. Mermelstein, "Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 28, no. 4, pp. 357–366 (1980 Aug.).

[20] The Mathworks Inc., "Audio Toolbox," (2022 Sep.) https://de.mathworks.com/products/audio.html.

[21] G. Peeters, "A Large Set of Audio Features for Sound Description (Similarity and Classification) in the CUIDADO Project," (2004 Apr.). http://recherche.ircam.fr/anasyn/peeters/ARTICLES/Peeters_2003_cuidadoaudiofeatures.pdf.

[22] H. Misra, S. Ikbal, H. Bourlard, and H. Hermansky, "Spectral Entropy Based Feature for Robust ASR," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (Montreal, Canada) (2004 May). https://doi.org/10.1109/icassp.2004.1325955.

[23] A. Pikrakis, T. Giannakopoulos, and S. Theodoridis, "A Speech/Music Discriminator of Radio Recordings Based on Dynamic Programming and Bayesian Networks", *IEEE Trans. Multimedia*, vol. 10, no. 5, pp. 846–857 (2008 Aug.). https://doi.org/10.1109/TMM.2008.922870.

[24] E. Scheirer and M. Slaney, "Construction and Evaluation of a Robust Multifeature Speech/Music Discriminator," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP*, vol. 2, pp. 1331–1334 (Munich, Germany) (1997 Apr.). https://doi.org/10.1109/ICASSP.1997.596192.

[25] S. Hatano and T. Hashimoto, "Booming Index as a Measure for Evaluating Booming Sensation," in *Proceedings of the 29th International Congress and Exhibition on Noise Control Engineering (INTER-NOISE)*, pp. 4332–4336 (Nice, France) (2000 Aug.).

[26] AudioCommons and Institute of Sound Recording, "Timbral Models," (2022 Apr.) https://www.audiocommons.org/ (accessed 2019).

[27] H. Fastl and E. Zwicker, *Psychoacoustics: Facts and Models*, 3rd ed. (Springer-Verlag Berlin Heidelberg, Berlin, Germany, 2007). https://doi.org/10.1007/978-3-540-68888-4.

[28] IEC, "Electroacoustics - Sound Level Meters - Part 1: Specifications," *International Standard 61672-1:2013Standard 61672-1:2013* (2013 Sep.).

[29] ITU-R, "Algorithms to Measure Audio Programme Loudness and True-Peak Audio Level," *Recommendation ITU-R BS.1770-4* (2015 Oct.).

[30] J. Blauert, *Spatial Hearing - The Psychophysics of Human Sound Localization*, 2nd ed. (The MIT Press, Cambridge, MA, 1997).

[31] P. Majdak, C. Hollomey, and R. Baumgartner, "AMT 1.x: A Toolbox for Reproducible Research in Auditory Modeling," *Acta Acust.*, vol. 6, paper 19 (2022 May). https://doi.org/10.1051/aacus/2022011.

[32] A. Andreopoulou and B. F. G. Katz, "Identification of Perceptually Relevant Methods of Inter-aural Time Difference Estimation," *J. Acoust. Soc. Am.*, vol. 142, no. 2, pp. 588–598 (2017 Aug.). https://doi.org/10.1121/1.4996457.

[33] V. Pulkki, "Spatial Sound Reproduction With Directional Audio Coding," *J. Audio Eng. Soc.*, vol. 55, no. 6, pp. 503–516 (2007 Jun.).

[34] A. Politis, *Microphone Array Processing for Parametric Spatial Audio Techniques*, Ph.D. thesis, Aalto University, Espoo, Finland (2016 Nov.).

[35] IEC, "Sound System Equipment – Part 5: Loudspeakers," *International Standard 60268-5:2003* (2003 May).

[36] D. Rudrich, S. Grill, M. Huber and IEM, "IEM Plug-in Suite v1.13.0," https://plugins.iem.at/ (Septmeber 15th 2022).

[37] C. Schörkhuber, M. Zaunschirm, and R. Höldrich, "Binaural Rendering of Ambisonic Signals via Magnitude Least Squares," in *Proceedings of the Fortschritte der Akustik - DAGA*, pp. 339–342 (Munich, Germany) (2018).

[38] B. Bernschütz, "A Spherical Far Field HRIR/HRTF Compilation of the Neumann KU 100," in *Proceedings of the Fortschritte der Akustik - AIA-DAGA*, pp. 592–595 (Merano, Italy) (2013).

[39] G. Elko and J. Meyer, "Eigenbeam Data: Specification for Eigenbeams," https://mhacoustics.com/sites/default/files/Eigenbeam Datasheet_R01A.pdf (2016).

[40] S. Brown and D. Sen, "Error Analysis of Spherical Harmonic Soundfield Representations in Terms of Truncation and Aliasing Errors," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP*, pp. 360–364 (Vancouver, Canada) (2013 May). https://doi.org/10.1109/ICASSP.2013.6637669.

[41] J. Bortz and C. Schuster, *Statistik für Human- und Sozialwissenschaftler* (Springer Berlin, Heidelberg, Berlin, Germany, 2010). https://doi.org/10.1007/978-3-642-12770-0.

[42] D. Barber, *Bayesian Reasoning and Machine Learning* (Cambridge University Press, Cambridge, UK, 2012). https://doi.org/10.1017/CBO9780511804779.

[43] F. Pedregosa, G. Varoquaux, A. Gramfort, et al., "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 12, no. 85, pp. 2825–2830 (2011 Oct.).

[44] G. J. Lautenschlager, C. E. Lance, and V. L. Flaherty, "Parallel Analysis Criteria: Revised Equations for Estimating the Latent Roots of Random Data Correlation Matrices," *Educ. Psychol. Meas.*, vol. 49, no. 2, pp. 339–345 (1989 Jun.). https://doi.org/10.1177/0013164489492006.

[45] A. Hyvärinen and E. Oja, "Independent Component Analysis: Algorithms and Applications," *Neural Networks*, vol. 13, nos. 4–5, pp. 411–430 (2000 Jun.). https://doi.org/10.1016/S0893-6080(00)00026-5.

[46] "Richard Wagner 3.0," https://richard-wagner3-0.de/en/index_en.html (Septmeber 15th 2022).

## THE AUTHORS


Jakob Bergner


Daphne Schössow


Stephan Preihs


Jürgen Peissig

Jakob Bergner received his M.Sc. in Audio Communication and Technology from TU Berlin in 2014. After his studies, he joined the Virtual Acoustics group at Fraunhofer IDMT and the Media Technology group at Technical University Ilmenau for project-oriented research on spatial audio technologies. In 2017, he joined the Institute of Communications Technology at Leibniz University Hannover as a research assistant where he received his doctoral degree (Dr.-Ing.) in 2023 for his work on describing general acoustic environments by means of acoustic dimensions for comparison and modeling purposes. He then moved back to Fraunhofer IDMT in 2023 as a researcher and project coordinator. His main research topics focus on audio signal processing, sound field analysis, psychoacoustic assessment, and soundscape studies.

•

Daphne Schössow received her M.Sc. degree in Technical Informatics in 2020 at Leibniz University Hanover with her thesis "Laboratory Study on Annoyance of Wind Turbine Noise by Employing Spatial Sound Field Reproduction." After participating as a graduate assistant in the research of the Institute of Communications Technology at Leibniz University Hanover, she joined the group in 2021 as a research assistant and Ph.D. student focusing on the mulitmodal perception of acoustic environments in laboratory scenarios.

•

Stephan Preihs is a postdoctoral researcher, coordinator, and group leader at the Institute of Communications Technology of Leibniz University Hannover since 2017. He received a doctoral degree (Dr.-Ing.) in electrical engineering from Leibniz University Hannover in 2016 for his contributions on low-delay, high-quality audio coding. His current work focuses on digital signal processing for audio and acoustics, immersive virtual and augmented reality audio as well as psychoacoustic analysis and modeling.

•

Jürgen Peissig received his Ph.D. in Physics from the University of Göttingen, Germany, in the fields of acoustics, psychoacoustics, and digital signal processing. In 1991, he worked at Bell Laboratories, Murray Hill, in the group of Gary Elko. In 1995, he joined Sennheiser Electronics R&D in Germany. Since 2004, he has been lecturing on acoustics and signal processing at Leibniz University of Hannover. Beginning in 2005, he was responsible for the Sennheiser research facility in Palo Alto, CA, focusing on digital signal processing for audio. Since 2014, Dr. Peissig is heading the communications systems group at the Institute of Communications Technology of Leibniz University of Hannover. Dr. Peissig is a member of the IEEE Communications Society, Audio Engineering Society, and German Acoustical Society.