



On the identification and assessment of underlying acoustic dimensions of soundscapes

Jakob Bergner*  and Jürgen Peissig 

Institute of Communications Technology, Leibniz University Hannover, Appelstr. 9A, 30167 Hannover, Germany

Received 6 July 2022, Accepted 14 September 2022

Abstract – The concept of soundscapes according to ISO 12913-1/-2/-3 proposes a descriptive framework based on a triangulation between the entities *acoustic environment*, *person* and *context*. While research on the person-related dimensions is well established, there is not yet complete agreement on the relevant indicators and dimensions for the pure description of acoustic environments. Therefore, this work attempts to identify acoustic dimensions that actually vary between different acoustic environments and thus can be used to characterize them. To this end, an exploratory, data-based approach was taken. A database of Ambisonics soundscape recordings (approx. 12.5 h) was first analyzed using a variety of signal-based acoustic indicators ($N_i = 326$) within the categories *loudness*, *quality*, *spaciousness* and *time*. Multivariate statistical methods were then applied to identify compound and interpretable acoustic dimensions. The interpretation of the results reveals 8 independent dimensions “Loudness”, “Directivity”, “Timbre”, “High-Frequency Timbre”, “Dynamic Range”, “High-Frequency Amplitude Modulation”, “Loudness Progression” and “Mid-High-Frequency Amplitude Modulation” to be statistically relevant. These derived latent acoustic dimensions explain 48.76% of the observed total variance and form a physical basis for the description of acoustic environments. Although all baseline indicators were selected for perceptual reasons, validation must be done through appropriate listening tests in future.

Keywords: Soundscape, Underlying acoustic dimensions, Statistical signal processing, Multivariate statistics

1 Introduction

Soundscape, as defined in ISO 12913-1 [1], is understood as a multidimensional framework whose evaluation is recommended by a triangulation between the aspects person, context and acoustic environment (AE) [2]. With the increase of information about each of these aspects, the description and interpretability of a specific soundscape gains validity. As soundscape itself is a multi-disciplinary concept, research in this area follows diverse approaches and paradigms. There is a considerable body of research on the human-centered approach of soundscape, for example the identification of fundamental emotional dimensions [3–5]. The dimensions for affective quality defined in [6], namely valence and arousal, were developed by means of factor or principle component analysis of a multitude of attribute scales [4, 7]. In contrast, there is only little agreement on the mere acoustical parameters, indicators and dimensions that describe the physical aspects of soundscapes in a discriminative way. Rather, the choice of parameters as indicators for individual hypotheses depends strongly on the respective discipline. Research attempts to model sound and noise quality and perception for example

rely often on A- and C-weighted sound pressure levels or conventional psychoacoustic measures such as loudness, sharpness, roughness and fluctuation [8]. An extensive review of models and underlying acoustic and psychoacoustic indicators can be found in [9]. At the same time, researchers agree that these and other known parameters are not sufficient to model emotional arousal and valence [5], also due to the lack of contextual and other non-acoustic indicators. A different application in computer science is the training of algorithms for extracting information from recordings of acoustic environments such as in acoustic event detection (AED), acoustic event classification (AEC), acoustic scene classification (ASC) or acoustic scene analysis (ASA). Here, parameters like short-time spectrograms or Mel-frequency cepstral coefficients (MFCCs) are widely used. An attempt to cluster acoustic events on the basis of selected acoustic features for the mapping of urban sounds can be found in [10]. However, the general question of which parameters and parameter combinations actually describe an acoustic environment adequately has not yet been answered satisfactorily. An exemplary overview of considered parameters for the respective research purposes can be found in Table 1. In the case of annoyance or quality modeling of soundscapes, an important question would be whether the selected signal properties are indeed capable

*Corresponding author: jakob.bergner@ikt.uni-hannover.de

Table 1. Exemplary overview on research using different sets of signal parameters.

| Author(s) | Purpose | Parameters |
|--------------------------------|-------------------------------------|--|
| ISO 2019 [6] | Soundscape assessment | L_{Aeq} , L_{Ceq} , loudness (ISO 532-1/-2), sharpness, roughness, fluctuation strength |
| Green and Murphy [24] | ASC | 20 MFCCs, FOA DirAC |
| Rey Gozalo et al. [29] | Soundscape perception modeling | L_{eq} , L_{Aeq} , loudness (ISO 532-1), sharpness |
| Jeon and Hong [30] | ASC, soundscape perception modeling | L_{Aeq} , L_{Ceq} , spectral centroid |
| Preis et al. [31] | Soundscape perception modeling | L_{Aeq} , loudness (ISO 532-1), sharpness, fluctuation strength, roughness |
| Park et al. [10] | AED, AEC, SIR | RMS, 13 MFCCs, zero-crossing rate, spectral centroid, flatness, flux, and spread |
| Cain et al. [4] | Soundscape perception modeling | L_{Aeq} , loudness (ISO 532-1), sharpness, roughness, fluctuation strength, intelligibility |
| Axelsson et al. [7] | Soundscape perception modeling | L_{Aeq} , L_{Ceq} , loudness (ISO 532-1) |
| Nilsson et al. [32] | Soundscape quality | L_{Aeq} , L_{Ceq} , loudness (ISO 532-1), spectral center of gravity, music-likeness, number of noise events |
| Persson Waye and Öhrström [33] | Annoyance modeling | Loudness (ISO 532-1), sharpness, tonality, roughness, fluctuation strength, modulation |

of sufficiently modeling the annoyance or whether, for example, a particular combination of parameters has unexpected effects. This leads to manifold approaches to find suitable parameters that can act as an appropriate indicator for perceptual, cognitive or emotional reactions. This work therefore aims to support this research direction by collecting potential acoustic parameters on the basis of which soundscapes with distinguishable human-related properties can be evaluated. It follows the approach that was conceptually presented previously by the authors in [11]. To this end, this paper takes a step back and examines which acoustic parameters occur at all with some variance in soundscapes. This exploratory approach is motivated by the development of the affective qualities, where a multitude of attributes are aggregated to a small number of emotional dimensions. In a similar way, a multitude of acoustic indicators are taken into account to form underlying, latent acoustic dimensions. With the aim that these dimensions are seen as distinguishing characteristics of acoustic environments, a variety of applications can be pursued such as modeling of human-centered responses [12, 13], the ecological validation of soundscape reproductions where certain acoustic properties are to be preserved [14, 15] or the training of algorithms for automatically deriving information from soundscape recordings [16].

2 Indicators for acoustic assessment

For the description of an acoustic environment, distinct quantifiable indicators must be identified and selected. Since the description aims to assess human-centered perception, a-priori categories for acoustic indicators are derived from semantic description: quality (in the following referred to as **Q**), loudness (**L**), spaciousness (**S**) and time (**T**). The category *quality* must not be confused with *valence* but

represents characteristics, that helps human beings to identify sound sources, such as information on timbre and spectral composition as well as short time temporal succession. *Loudness* distinguishes whether an acoustic environment is perceived as loud or soft in volume, *spaciousness* represents location, distribution and envelopment of both the sound sources and indistinguishable background noise. The category *time* describes how the acoustic scene changes over time. Suitable indicators from literature sources (Tab. 1 and others) are assigned to these categories in the following listing. A detailed description, scientific sources and implementation of each indicator can be found in the [Supplementary Material A](#).

Quality: MFCC, Spectral Brightness, Spec. Centroid, Spec. Crest Factor, Spec. Decrease, Spec. Entropy, Spec. Flatness, Spec. Flux, Spec. Irregularity, Spec. Kurtosis, Spec. Roll-Off, Spec. Skewness, Spec. Spread, Timbral Booming, Roughness, Sharpness, Fluctuation Strength.

Loudness: SPL (A-/Z-weighting), Octave Band Energy, Loudness (ISO 532-1/2), LUFS (EBU R 128).

Spaciousness: ILD, ITD, IACC, IC, Direction of Arrival (hor., vert.), Diffuseness, Directivity Index (hor., vert., sph.), Ambisonics Energy Ratio.

Time: amplitude modulation (frequency and depth; periodic and stochastic); time series of all above indicators.

For this work, each indicator is calculated as time series for overlapping frames of 100 ms each and hop size of 50 ms to respect both time-integrating behavior of the human auditory model [17] and time-variance of acoustic scenes. It is recognized here that there may be acoustic events and psychoacoustic effects that are difficult to detect with this temporal resolution. At the same time, averaging through large analysis windows contributes to the increased robustness of the results against statistical and measurement noise. Furthermore, the majority of indicators is calculated frequency-dependent. For that, the broadband

Table 2. Frequency limits in Hz of analysis bands.

| ID | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|-------|----|----|-----|-----|-----|------|------|------|------|--------|
| f_c | 31 | 62 | 125 | 250 | 500 | 1000 | 2000 | 4000 | 8000 | 16,000 |

signals are filtered using 10 octave filters with center frequencies given in [Table 2](#) and indicators are calculated for each octave band individually. Again, this spectral resolution is not sufficient to separate the filter bands of human hearing or the spectral composition of individual sound sources. However, it offers the possibility to detect a general and interpretable frequency dependence of the acoustic indicators. The indicators themselves are based on one of three signal representations of the same acoustic environment: The quality and loudness indicators may be calculated either from a monophonic pressure representation or from a binaural signal, while the spaciousness indicators require binaural and spherical harmonic (Ambisonics) signal representation of the three-dimensional soundfield. The latter two representations incorporate spatial information of an acoustic environment such as the location of sound sources or the envelopment of sound. In order to maintain consistency and to reduce data complexity, all three representations stem from the same recording of a specific acoustic environment. For that, microphone array recordings are necessary that can be transformed into the spherical harmonic domain as it is established in Ambisonics encoding and rendering [18]. The order of the ambisonic recordings generally determine the spatial confidence. However, even first-order Ambisonics (FOA) recordings are suitable for the analysis in this work. The binaural representation is derived by convolution with appropriate head-related transfer functions (HRTF) [19] as it is established in [20, 21]. The monophonic sound pressure representation on the other hand side is proportional to the 0th-order Ambisonics component [22].

3 Determining underlying dimensions

The idea pursued in this work is that the multitude of indicators contain information describing the properties of an acoustic environment that are relevant when a human being perceives and contextualises the same environment. Just as humans can classify their environment acoustically on the basis of their two ear signals, a procedure is now to be developed here that provides an abstract construct for the description and identification of acoustic environments on the basis of the indicators presented in the previous section. In other words it is assumed that the observed indicators above are realizations of certain underlying acoustic dimensions that characterize an acoustic scene or environment. These assumptions allow the application of exploratory factor analysis (FA) as schematically depicted in [Figure 1](#). Similar to the related principle component analysis (PCA), FA can be used here to aggregate data variances (and thus information) by transforming the observed indicator time series from an original space into an optimized

space of latent dimensions. The methodological differences between PCA and FA concern the perspective: while PCA assumes that the observed indicators constitute the ground truth, which in turn can be described by principal components, FA implies that the (hidden) latent factors constitute the ground truth and the observed indicators are more or less arbitrary realizations of it. For the sake of comparability, some taxonomies, measures and results of PCA are placed alongside those of FA in the following. The operation itself to obtain the factor scores \mathbf{Y} in the optimized space is realized by matrix multiplication as shown in equation (1)

$$\mathbf{Y} = \mathbf{X} \cdot \mathbf{L}, \quad (1)$$

where \mathbf{X} is a $[N_o \times N_i]$ matrix (N_o : number of observations; N_i : number of indicators) of the original data and \mathbf{L} a specific loading matrix of dimension $[N_i \times N_f]$ (N_f : number of factors). The loading matrix comprises the individual weights of each indicator into each factor. The sum down the rows, i.e. among indicators yields the sum of square loadings (in PCA: eigenvalues of covariance matrix) or explained variance of a certain factor

$$s_j^2 = \sum_{i=1}^{N_i} l_{ij}^2. \quad (2)$$

This measure indicates the weight of a particular j th factor, which is important when deciding which factors to retain. Dividing \mathbf{L} by the respective explained variances yields the relative Loading \mathbf{L}_{rel} (in PCA: eigenvectors of covariance) that includes the assignment of the indicators to the respective factors and represents the direction of the transformation:

$$\mathbf{L}_{\text{rel}} = \mathbf{L} \cdot \text{diag}\{\mathbf{s}\}^{-1}. \quad (3)$$

In contrast to PCA, the factors in FA only express the common part of variance of the observed indicators. That means in practice that each indicator may inhibit portions of specific variance ϵ_s as well as measurement noise ϵ_n which both is not included in the factors as denoted in [Figure 1](#) with $\epsilon_i = \epsilon_{si} + \epsilon_{ni}$. Hence, we allow the indicators to be imperfect realizations of the factors which relaxes the necessary requirements of the indicators.

In order to apply FA to indicators of different scales and units, preprocessing of the initial indicator vectors must be applied. For that, an interval range of expected values was defined for each indicator and scaling was applied accordingly to derive relative values within this interval. Since FA is only capable to identify linear relationships, non-linear indicators must also be treated accordingly. Ratio-scaled indicators with reference to frequency/Hz are converted to frequency in octaves relative to 10 Hz to regard the logarithmic behavior of auditory pitch perception. Conversions and expectation intervals for each indicator can be taken from the [Supplementary Material A](#). Finally, a z -standardization was applied to each indicator, that means removal of the mean and normalization to unit variance.

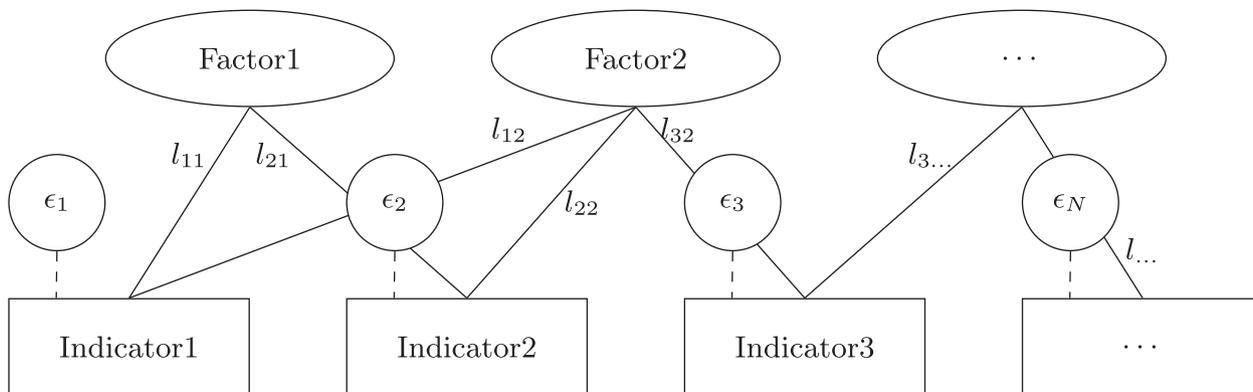


Figure 1. Concept of factor analysis with loadings l_{ij} and unique variances ϵ_i .

Pure FA produces mutually independent (uncorrelated) factors where the first factor includes maximum variance. However this might result in a loading matrix that is difficult to interpret. In these cases, a further rotation of the loading matrix \mathbf{L} aims for a simple structure with few high loadings and many low loadings. In this work the orthogonal rotation method *varimax* was chosen to preserve uncorrelated factors while increasing interpretability.

4 Application

The following section describes how the previously presented methodology is applied to real observational data of acoustic environments. To make generalizable statements about underlying dimensions, the sample draw of AE recordings for developing the final loading matrix \mathbf{L} must be chosen with care. This work focuses on acoustic environments that can potentially be part of soundscape research and consists of indoor and outdoor recordings of public places with and without human impact. The selection consists of publicly available recording data bases where Ambisonics or microphone array recordings are utilised. The listed databases in the following are used.

ARTE [23]: 13 mixed-order Ambisonics (4th/7th-order) recordings between 01:21 and 02:30 min: library, office, church, living room, café, dinner party, street, train station and food court.

Eigenscape [24]: 8×8 4th-order Ambisonics recordings of 10 min each: beach, street, park, pedestrian zone, shopping centre, train station, woodland.

Soundfield by Røde Ambisonic Sound Library [25]: selection of 35 first-order Ambisonics recordings between 00:07 and 06:41 min: indoor crowd, playground, car, foyer, library, mall, market, metro, street, steam train, traffic and train station.

All in all $N_o = 903,735$ observations of 100 ms were analyzed for $N_i = 326$ indicators. The multivariate methods PCA and FA both with and without subsequent *varimax* rotation were applied and performance metrics were compared. Figure 2 shows the cumulative explained variance portions of the four methods. In order to identify the most relevant factors or principle components respectively,

parallel analysis was applied as well as the Kaiser criterion. The result of this relevance analysis can be found in Table 3. We recall that the goal of the multivariate methods here is to find latent dimensions that are manageable in number and interpretable at the same time. Thus we not only evaluate the mere number of relevant components but also their composition, i.e. what indicators contribute considerably to a certain dimension. We find that the *varimax* rotation leads to a fewer number of indicators with higher loading that contribute to the latent dimensions as intended, which is why we choose this processing step to be beneficial. We can also observe that the parallel analysis assumes a lower number of components to be relevant which also aligns with expectations as well as our aim. After all, the first 8 *varimax*-rotated factors explain 49% of the total variance, which is generally just a moderate result but a reasonable starting point given a large input of $N_o = 903,735$ observations with $N_i = 326$ indicators. An investigation of the composition of these 8 factors reveals its interpretability in terms of acoustical semantics. Table 4 lists the indicators that contribute to a certain factor. The indicators are sorted in descending order by their absolute relative loading in parentheses. Only those indicators are listed whose cumulative sum of squares describes at least half of the respective factor's variance and thus characterize it.

5 Results

The factors are sorted with decreasing amount of explained variance. It has to be mentioned that the amount of variance that can be explained with a certain factor drastically depends on the initial choice of indicators. From that follows that the explanatory power of the absolute amount of explained variance within the factors should not be overestimated rather than the indicator composition itself.

Factor 1 obviously describes the level or loudness of the soundscape recording. Indicators such as loudness (Zwicker), A-weighted SPL and loudness (LUFS) dominate this factor especially within the frequency bands number 3–5 (corr. to $f_c = 250 \dots 1000$ Hz).

Factor 3, which explains the second most portion of variance, comprises the spherical directivity indices of the

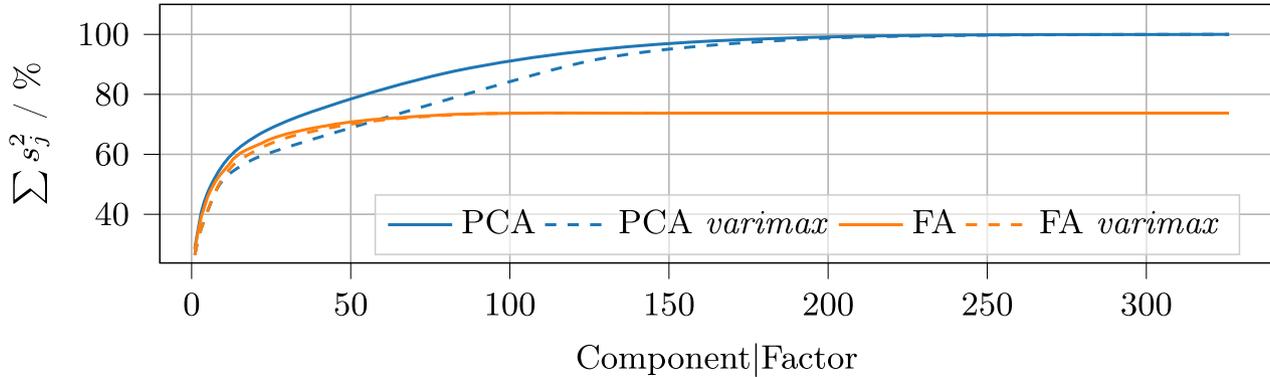


Figure 2. Explained variance ratio in % for PCA and FA both with and without *varimax* rotation.

Table 3. Number $N_{f,r}$ of relevant factors (FA) or principle components (PCA) according to Kaiser criterion and parallel analysis.

| | Kaiser criterion | | Parallel analysis | |
|--------------------|------------------|------------------------------|-------------------|------------------------------|
| | $N_{f,r}$ | $\sum_{j=1}^{N_{f,r}} s_j^2$ | $N_{f,r}$ | $\sum_{j=1}^{N_{f,r}} s_j^2$ |
| PCA | 59 | 265.21 (81.35%) | 8 | 174.10 (53.41%) |
| PCA <i>varimax</i> | 85 | 259.72 (79.69%) | 8 | 158.35 (48.57%) |
| FA | 28 | 216.06 (66.28%) | 8 | 168.76 (51.77%) |
| FA <i>varimax</i> | 30 | 213.74 (66.56%) | 8 | 158.97 (48.76%) |

Table 4. Indicator composition of the 8 most relevant factors. Trailing index numbers denote the respective frequency band ID as listed in Table 2. Factors in parentheses indicate respective loadings.

| Factor j | s_j^2 | Indicators |
|------------|---------------|--|
| 1 | 86.10 (26.4%) | loudnessZwicker _(0.106) , loudnessZwickerBands04 _(0.105) , loudnessZwickerBands05 _(0.105) , loudnessZwickerBands03 _(0.105) , LA _(0.105) , LAeq _(0.105) , LABands03 _(0.104) , LAeqBands03 _(0.104) , LAeqBands05 _(0.104) , LABands05 _(0.104) , lufsPeakBands03 _(0.104) , LABands04 _(0.104) , LAeqBands04 _(0.104) , lufsPeakBands04 _(0.104) , lufsPeakBands05 _(0.104) , oct06 _(0.104) , LApeakBands03 _(0.104) , oct04 _(0.104) , LAmxBands03 _(0.104) , oct05 _(0.104) , LAmxBands04 _(0.103) , LAmx _(0.103) , loudnessZwickerBands06 _(0.102) , LAmxBands05 _(0.102) , LApeakBands04 _(0.102) , mfcc00 _(0.102) , lufsPeak _(0.102) , LApeakBands02 _(0.102) , LAeqBands02 _(0.102) , LABands02 _(0.102) , LApeakBands05 _(0.101) , LAmxBands02 _(0.101) , loudnessZwickerBands02 _(0.101) , LApeak _(0.101) , oct07 _(0.101) , LABands06 _(0.101) , LAeqBands06 _(0.101) , lufsPeakBands06 _(0.100) , lufsPeakBands02 _(0.099) , lufsMomBands03 _(0.098) , lufsMomBands04 _(0.098) , oct03 _(0.098) , lufsMomBands05 _(0.097) , LAmxBands06 _(0.097) , LApeakBands06 _(0.096) , loudnessZwickerBands07 _(0.095) , LApeakBands01 _(0.094) , lufsMom _(0.094) , lufsMomBands06 _(0.093) |
| 3 | 16.79 (5.2%) | sphDI06 _(-0.225) , sphDI07 _(-0.225) , sphDI05 _(-0.223) , sphDIEl07 _(-0.222) , sphDI08 _(-0.216) , sphDIEl05 _(-0.216) , sphDIEl08 _(-0.214) , sphDIEl04 _(-0.208) , sphDI04 _(-0.207) , sphDIEl06 _(0.198) , sphDIAz06 _(-0.194) |
| 2 | 11.95 (3.7%) | spectralCentroid _(0.247) , spectralRolloffPoint _(0.243) , spectralEntropy _(0.242) , spectralCrest _(-0.214) , spectralSkewness _(-0.194) , spectralSpread _(0.183) , spectralFlatness _(0.173) , oct01 _(-0.170) , lufsPeakBands00 _(-0.168) , spectralKurtosis _(-0.167) , oct02 _(0.162) , LABands00 _(-0.154) , lufsPeakBands01 _(-0.153) , LAeqBands00 _(-0.151) |
| 5 | 10.27 (3.2%) | mfcc01 _(-0.244) , sharp _(0.220) , lufsPeakBands09 _(0.189) , LABands09 _(0.187) , LAmxBands09 _(0.185) , LAeqBands09 _(0.185) , LAmxBands08 _(0.182) , LABands08 _(0.180) , LAeqBands08 _(0.178) , lufsPeakBands08 _(0.177) , lufsMomBands09 _(0.174) , lufsMomBands08 _(0.167) , LApeakBands08 _(0.163) , LApeakBands09 _(0.162) , oct09 _(0.162) |
| 7 | 9.17 (2.8%) | lufsRange _(-0.311) , lufsRangeBands04 _(-0.303) , lufsRangeBands03 _(-0.299) , lufsRangeBands05 _(-0.293) , lufsRangeBands02 _(-0.283) , lufsRangeBands01 _(-0.274) |
| 6 | 8.84 (2.7%) | modDepthS09 _(0.303) , modDepthS08 _(0.292) , modDepthP109 _(0.291) , modDepthP209 _(0.287) , modDepthP309 _(0.287) , modDepthP308 _(0.262) , modDepthP208 _(0.261) |
| 4 | 8.28 (2.5%) | lufsShort _(-0.295) , lufsShortBands09 _(-0.292) , lufsShortBands07 _(-0.290) , lufsShortBands08 _(-0.288) , lufsShortBands06 _(-0.284) , lufsShortBands01 _(-0.284) |
| 8 | 7.58 (2.3%) | modDepthS06 _(0.276) , modDepthP106 _(0.273) , modDepthP105 _(0.261) , modDepthS05 _(0.255) , modDepthP306 _(0.250) , modDepthP206 _(0.247) , modDepthP305 _(0.230) , modDepthP205 _(0.227) |

Table 5. Summarized semantic descriptors for the first eight relevant factors.

| Factor j | s_j^2 | Semantic descriptor |
|------------|---------------|---|
| 1 | 86.10 (26.4%) | “Loudness” |
| 3 | 16.79 (5.2%) | “Directivity” |
| 2 | 11.95 (3.7%) | “Timbre” |
| 5 | 10.27 (3.2%) | “High-Frequency Timbre” |
| 7 | 9.17 (2.8%) | “Dynamic Range” |
| 6 | 8.84 (2.7%) | “High-Frequency Amplitude Modulation” |
| 4 | 8.28 (2.5%) | “Loudness Progression” |
| 8 | 7.58 (2.3%) | “Mid-High-Frequency Amplitude Modulation” |

incoming soundfield representing information whether the sound energy arrives from a certain direction or region or if it surrounds the respective receiver position. Again the mid-high frequency bands are prominent in discriminating the observations.

Factor 2 includes mainly spectral characteristics. To have these indicators forming a prominent factor is somewhat surprising from a statistical point of view since these indicators are not calculated in frequency bands and thus contribute only once to the total explained variance. It is also interesting that these indicators form a factor together with SPL and loudness indicators in the low-frequency bands 0 and 1.

Factor 5 includes high frequency content such as SPL and loudness in the frequency bands 8 and 9 ($f_c = 8 \dots 16$ kHz) but also sharpness which is measure for high frequency spectral characteristic. Obviously this higher frequency region is more or less independent of the general spectral timbre as described in the previous factor 2.

Factor 7 comprises loudness range indicators following the algorithms according to [26–28]. It has to be mentioned that the calculation of this loudness range is usually performed as singular value for a specific audio content (e.g. music or movie). The calculation as time series of soundscape recordings describes the loudness range of the respective previous time interval from start of the recording to the current timestamp. This leads to the fact that the observations depend on previous time periods and thus violate the requirements for FA. This behavior is also reflected statistically, as no other indicator (group) contributes to factor 7. Therefore, it may be necessary to omit this factor altogether from further analysis.

A different analysis of temporal behaviour of loudness or level can be observed in factor 6 which includes especially the modulation depth of the first three dominant periodic modulations as (modDepthP) as well as the remaining stochastic modulation (modDepthS) for the high frequency range of bands 8 and 9 ($f_c = 8 \dots 16$ kHz). This modulation behaviour obviously differs from that in the mid-high range (bands 5 and 6) that we can observe in factor 8.

Factor 4 mainly summarizes loudness indicators of short-time averaged LUFs. Since these indicators inhibit a moving time window of 3 s, the resulting factor can be interpreted as loudness progression. It is noteworthy that this specific temporal characteristic is statistically independent from general loudness in factor 1.

Table 6. Sample draw of soundscape excerpts for exemplary pairwise comparison.

| ID | Database | Name | Excerpt |
|----|------------|------------------------|-------------|
| 1 | ARTE | 01_Library | 01:30–02:00 |
| 2 | ARTE | 02_Office | 00:35–01:05 |
| 3 | ARTE | 04_Living_Room | 00:58–01:28 |
| 4 | ARTE | 07_Cafe_1 | 01:28–01:58 |
| 5 | ARTE | 09_Dinner_Party | 01:36–02:06 |
| 6 | ARTE | 11_Train_Station | 00:18–00:48 |
| 7 | ARTE | 12_Food_Court_1 | 00:36–01:06 |
| 8 | Eigenscape | Beach.7 | 05:38–06:08 |
| 9 | Eigenscape | Park.5 | 05:28–05:58 |
| 10 | Eigenscape | PedestrianZone.3 | 08:00–08:30 |
| 11 | Eigenscape | PedestrianZone.5 | 06:48–07:18 |
| 12 | Eigenscape | ShoppingCentre.8 | 05:24–05:54 |
| 13 | Eigenscape | TrainStation.6 | 03:24–03:54 |
| 14 | Eigenscape | Woodland.2 | 03:16–03:46 |
| 15 | Soundfield | Kids Playground 1 | 00:47–01:17 |
| 16 | Soundfield | Rural Market Busker | 01:40–02:10 |
| 17 | Soundfield | Steamtrain Exterior | 02:00–02:30 |
| 18 | Soundfield | St Kilda Road Traffic | 02:56–03:26 |
| 19 | Soundfield | Southern Cross Station | 00:30–01:00 |

In summary we observe that the first 8 relevant factors can be interpreted quite well in terms of dominant indicators that contribute to them. Table 5 lists the semantic descriptors that are proposed from the findings above.

6 Plausibility considerations

In order to conduct steps for validation of the identified acoustic dimensions, we show a comparative example at this point. This includes a sample draw of 19 excerpts of 30 seconds each of the soundscape recording data base as listed in Table 6. The sample includes examples from each recording database for a range of acoustic environment classes that potentially contain the three sound source classes according to ISO 12913-2, namely sounds of technology, sounds of nature and sounds of human beings. The excerpts were selected subjectively by the authors aiming for two criteria: (i) a homogeneous listening impression throughout a sample, to allow semantic and statistic description that is valid for the entire excerpt and (ii) samples that have potentially similarities and differences within

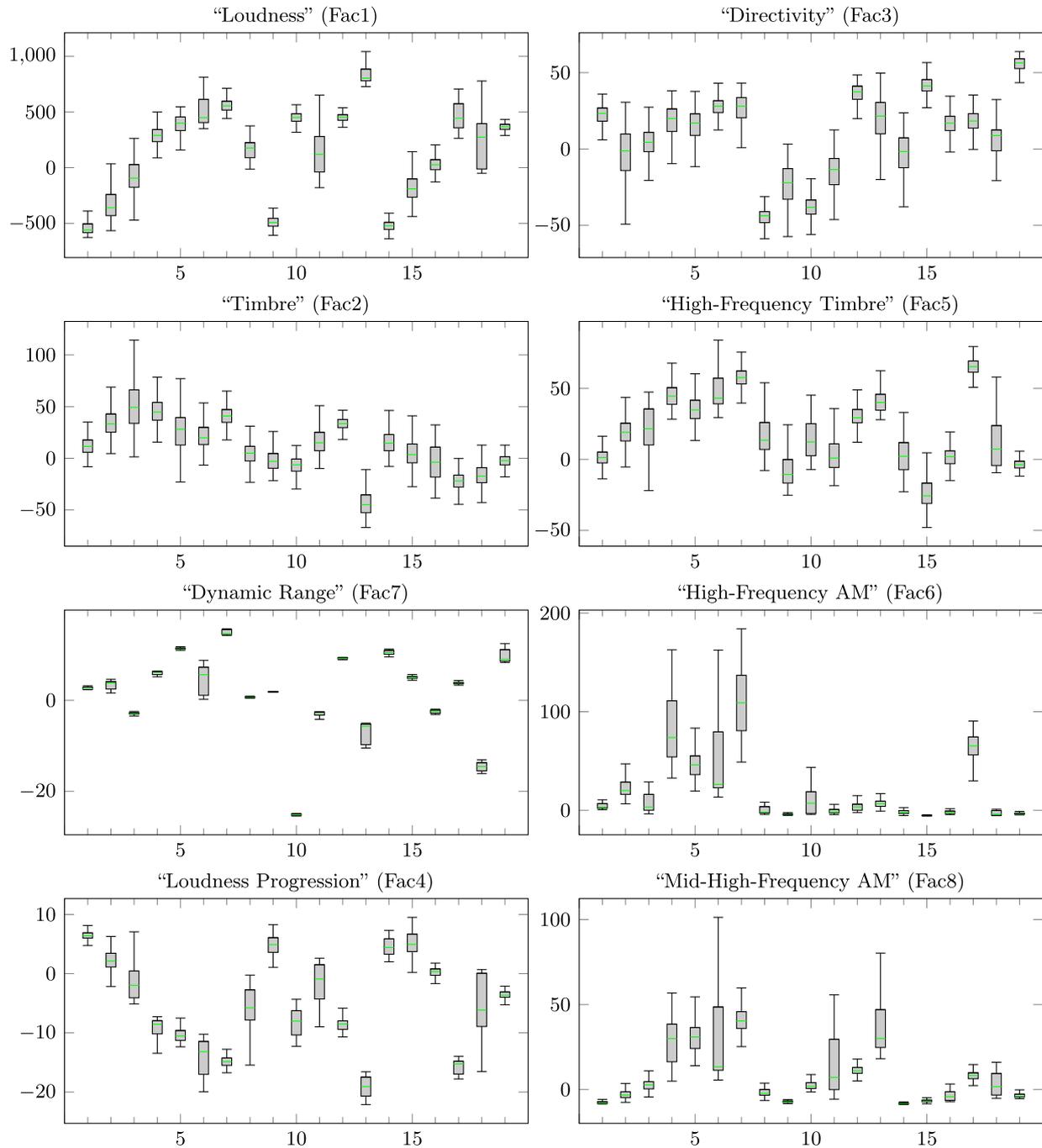


Figure 3. Distribution of factor scores amongst the soundscape recordings for each relevant factor.

the identified acoustic features. The acoustic indicators were calculated and extracted and the resulting factor scores according to equation (1) were deduced. The distributions of factor scores for these samples can be found in Figure 3 as boxplots of median, 25% and 75% quantiles where outliers are omitted for better visibility. For comparison reasons, the distribution of all 109 analyzed soundscape recordings are listed in the Supplementary Material B.

The samples' factor score distribution were analyzed with regards to normality and homoscedasticity which could not be asserted for every case and which is why

non-parametric statistic methods have been applied. A Kruskal–Wallis test on ranks proposes significant differences among the samples within all acoustic dimensions ($H > 9500$, $p < 0.01$). Subsequently, pairwise Dunn's posthoc tests with Bonferroni adjustment were performed comparing all 19 samples with each other for each dimension. The result whether each comparison pair differs significantly can be found in Figure 4. It is noteworthy that the majority of these comparisons exhibit strong significant differences with $p < 0.01$ (**, red tiles). This result might be influenced by the relatively large number of observations

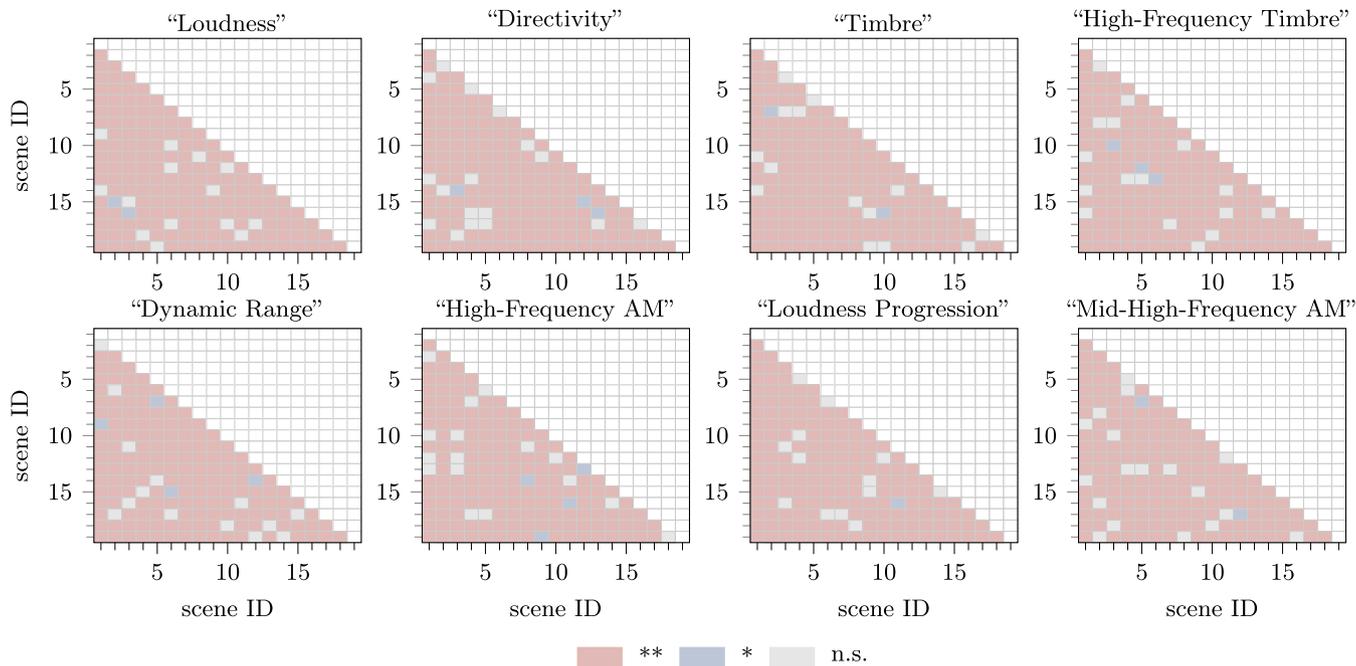


Figure 4. Statistical differences between sample soundscapes. Red: strong significance ($p < 0.01$), blue: moderate significance ($p < 0.05$), gray: no significance ($p > 0.05$).

(30 s \times 20 observations/s) and should at this point only describe the difference from a statistical point of view. An exemplary comparison of those soundscape excerpts that are specifically quiet in loudness, namely the library (ID: 1), park (ID: 9) and woodland (ID: 14) scenarios, shows similar and different properties with respect to their acoustic dimensions. The distributions of the dimension Loudness are relatively low (cf. Fig. 3, top left) and the differences between the soundscape excerpts are not significant (cf. Fig. 4, top left). The other acoustic dimensions of these excerpts show similar but still significantly different distributions. This underlines the expected discriminating characteristics of the observed soundscapes: even though the loudness dimension seems to be similar, other dimensions show significant differences.

7 Discussion

With the presented application of multivariate methods on a wide range of acoustic indicators of soundscape recordings, it is possible to extract statistically independent factors to serve as underlying acoustic dimensions. It could also be shown that an interpretation of these factors based on the indicator composition was feasible in terms of finding appropriate semantic descriptors. These descriptors can generally be assigned to the a-priori categories *loudness*, *quality*, *spatiousness* and *time* and thus confirm the assumption that acoustic environments can be described with these terms. The fact that each of these categories is represented by more than one factor (e.g. loudness: factor 1 and 4; quality/timbre: factor 2 and 5; time/modulation: factor 6 and 8) can be interpreted such that the selection

of indicators is crucial when physical characteristics of acoustic environment shall be described. With other words, if different acoustic indicators are chosen for the multivariate analysis, different factor compositions may be observed. In order to validate the deduced underlying acoustic dimensions, discriminative investigations must be rolled out in future. These include both the statistical differentiation of specific soundscape recordings as well as perceptual evaluation if these dimensions are also taken into account when human subjects characterize acoustic environments. Therefore, this paper should serve as an invitation to evaluate and refine the proposed acoustic dimensions.

8 Conclusion

The presented paper discusses the need of suitable acoustic descriptors for characterizing the physical properties of soundscapes. For that, an approach was pursued that is comparable to the identification of semantic dimensions of perceptual assessment, namely the application of multivariate statistic methods to reveal underlying constructs of observable variables. In this work these methods were adapted to acoustic signal indicators. In total 903,735 short-term observations of 326 indicators within 109 recordings of soundscapes were fed into factor analysis (FA) and relevant factors were deduced. With this set of eight underlying dimensions 49% of the overall observed variance could be explained and interpretable semantic descriptors could be found. The presented approach allows the description of acoustic environment in an efficient and comprehensive way. Various areas of application may benefit from this descriptive set of acoustic dimensions, e.g. computer-based

applications such as acoustic scene analysis and classification or perception-based applications such as soundscape quality estimation or annoyance modelling. Furthermore, if this approach receives confirmation, it can contribute to be used as a comparative benchmark method for soundscape description and analysis. The derived results are limited by certain implications and conditions. First, the initial choice of indicators may influence the statistic outcome especially within the amount of explained variance. Second, a perceptual validation by means of appropriate listening tests is still pending. Appropriate listening tests are currently conducted and results will be published in the near future. Third, the influence of signal analysis parametrization, including time and frequency resolution, Ambisonics decoding scheme, and binaural convolution, needs to be quantified, especially when merging statistical and perceptual results.

Conflict of interest

The authors declare no conflict of interest.

Data availability statement

Data are available on request from the authors.

Acknowledgments

The authors thank for the funding of two research projects in whose background the investigations on acoustic soundscape dimension could grow. First, *WEA-Akzeptanz* dealing with the modelling of wind turbine noise and its perception, funded by the German Federal Ministry of Economics and Energy (FKZ 0324134A) and second, *Richard Wagner 3.0* on perception of music reproduction, funded by “Niedersächsisches Vorab” (ZN3497).

Supplementary material

The supplementary material of this article is available at <https://acta-acustica.edpsciences.org/10.1051/aacus/2022042/olm>.

Supplementary material A: Indicator descriptions.

Supplementary material B: Distribution of factor scores amongst the soundscape recordings for each relevant factor.

References

1. ISO 12913-1: Acoustics soundscape. Part 1. Definition and conceptual framework. ISO, 2018.
2. J. Kang, B. Schulte-Fortkamp: Soundscape and the built environment. CRC Press, Boca Raton, 2016. <https://doi.org/10.1201/b19145>. ISBN 9781482226324.
3. A. Fiebig, P. Jordan, C.C. Moshona: Assessments of acoustic environments by emotions – the application of emotion theory in soundscape. *Frontiers in Psychology* 11 (2020). <https://doi.org/10.3389/fpsyg.2020.573041>. ISSN 16641078
4. R. Cain, P. Jennings, J. Poxon: The development and application of the emotional dimensions of a soundscape. *Applied Acoustics* 74, 2 (2013) 232–239. <https://doi.org/10.1016/j.apacoust.2011.11.006>. ISSN 0003-682X
5. F. Aletta, J. Kang, Ö. Axelsson: Soundscape descriptors and a conceptual framework for developing predictive soundscape models. *Landscape and Urban Planning* 149 (2016) 65–74. <https://doi.org/10.1016/j.landurbplan.2016.02.001>. ISSN 0169-2046
6. ISO 12913-2: Acoustics. Soundscape. Part 2. Data Collection and reporting requirements. ISO, 2019.
7. Ö. Axelsson, M.E. Nilsson, B. Berglund: A principal components model of soundscape perception. *Journal of the Acoustical Society of America* 128, 5 (2010) 2836–2846. <https://doi.org/10.1121/1.3493436>. ISSN 0001-4966
8. M.S. Engel, A. Fiebig, C. Pfaffenbach, J. Fels: A review of the use of psychoacoustic indicators on soundscape studies. *Current Pollution Reports* 7 (2021) 359–378. <https://doi.org/10.1007/s40726-021-00197-1>.
9. M. Lionello, F. Aletta, J. Kang: A systematic review of prediction models for the experience of urban soundscapes. *Applied Acoustics* 170 (2020). <https://doi.org/10.1016/j.apacoust.2020.107479>.
10. T.H. Park, J.H. Lee, J. You, M.J. Yoo, J. Turner: Towards soundscape information retrieval (SIR), in *Proceedings of the 11th Sound and Music Computing Conference, SMC, 14–20 September 2014, Athens, Greece*, pp. 1218–1225. ISBN 9789604661374
11. J. Bergner, S. Preihs, J. Peissig: Soundscape fingerprinting – methods and parameters for acoustic assessment, in *Fortschritte der Akustik – DAGA, 15–18 August 2021, Vienna, Austria*.
12. S. Preihs, J. Bergner, D. Schössow, J. Peissig: On predicting the perceived annoyance of wind turbine sound, in *Fortschritte der Akustik – DAGA, 15–18 August 2021, Vienna, Austria*.
13. J. Bergner, D. Schössow, S. Preihs, Y. Wycisk, K. Sander, R. Kopiecz, F. Platz: Analyzing the degree of immersion of music reproduction by means of acoustic fingerprinting, in *Fortschritte der Akustik – DAGA, 21–24 March 2022, Stuttgart, Germany*.
14. C. Guastavino, B.F.G. Katz, J.-D. Polack, D.J. Levitin, D. Dubois: Ecological validity of soundscape reproduction. *Acta Acustica United with Acustica* 91 (2004) 333–341.
15. J. Bergner, S. Preihs, J. Peissig: Investigation on ecological validity within higher order ambisonics reproductions of wind turbine noises, in *Proc. of Internoise, 16–19 June 2019, Madrid, Spain*.
16. N. Poschadel, C. Gill, S. Preihs, J. Peissig: CNN-based multi-class multi-label classification of sound scenes in the context of wind turbine sound emission measurements, in *Proc. of Internoise, 1–5 August 2021, Washington, DC, USA*.
17. J. Blauert: *Spatial hearing*. The MIT Press, 1997. ISBN 0-262-02413-6
18. B. Rafaely: *Fundamentals of spherical array processing*. Springer, Berlin, Heidelberg, 2015. <https://doi.org/10.1007/978-3-662-45664-4>. ISBN 978-3-662-45663-7
19. B. Bernschütz: A spherical far field HRIR/HRTF compilation of the Neumann KU 100, in *Fortschritte der Akustik – AIA-DAGA, 1821 March 2013, Merano, Italy*, pp. 592–595. http://www.audiogroup.web.fh-koeln.de/FILES/AIA-DAGA2013_HRIRs.pdf
20. D. Rudrich: *IEM Plug-in Suite*. IEM, 2021. <https://plugins.iem.at/>
21. C. Schörkhuber, M. Zaunschirm, R. Höldrich: Binaural rendering of Ambisonic signals via magnitude least squares, in *Fortschritte der Akustik – DAGA, 19–22 March 2018, Munich, Germany*, pp. 339–342.

22. V. Pulkki: Spatial sound reproduction with directional audio coding. *Journal of the Audio Engineering Society* 55, 6 (2007) 503–516.
23. A. Weisser, J.M. Buchholz, C. Oreinos, J. Badajoz-Davila, J. Galloway, T. Beechey, G. Keidser: The ambisonic recordings of typical environments (ARTE) database. *Acta Acustica United with Acustica* 105 (2019) 695–713. <https://doi.org/10.3813/AAA.919349>.
24. M. Green, D. Murphy: Eigenscape: a database of spatial acoustic scene recordings. *Applied Sciences* 7, 12 (2017) 1204. <https://doi.org/10.3390/app7111204>. ISSN 2076-3417
25. Soundfield by Røde: Ambisonic sound library, 2022. <https://library.soundfield.com/>
26. EBU-R 128: Loudness normalisation and permitted maximum level of audio signals, EBU, 2014.
27. ITU-R BS.1770-4: Algorithms to measure audio programme loudness and true-peak audio level. ITU, 2015.
28. EBU Tech 3342: Loudness range: a measure to supplement EBU R 128 loudness normalization. EBU, 2016.
29. G. Rey Gozalo, J. Trujillo Carmona, J.M. Barrigón Morillas, R. Vélchez-Gómez, V. Gómez Escobar: Relationship between objective acoustic indices and subjective assessments for the quality of soundscapes. *Applied Acoustics* 97 (2015) 1–10. <https://doi.org/10.1016/j.apacoust.2015.03.020>. ISSN 0003-682X
30. J.Y. Jeon, J.Y. Hong: Classification of urban park soundscapes through perceptions of the acoustical environments. *Landscape and Urban Planning* 141 (2015) 100–111. <https://doi.org/10.1016/j.landurbplan.2015.05.005>. ISSN 0169-2046
31. A. Preis, J. Kociński, H. Hafke-Dys, M. Wrzosek: Audio-visual interactions in environment assessment. *Science of the Total Environment* 523 (2015) 191–200. <https://doi.org/10.1016/j.scitotenv.2015.03.128>. ISSN 18791026
32. M.E. Nilsson, D. Botteldooren, B. De Coensel: Acoustic indicators of soundscape quality and noise annoyance in outdoor urban areas, in 19th International Congress on Acoustics, ICA, 2–7 September 2007, Madrid, Spain.
33. K. Persson Waye, E. Öhrström: Psycho-acoustic characters of relevance for annoyance of wind turbine noise. *Journal of Sound and Vibration* 250, 1 (2002) 65–73. <https://doi.org/10.1006/jsvi.2001.3905>. ISSN 0022-460X

Cite this article as: Bergner J. & Peissig J. 2022. On the identification and assessment of underlying acoustic dimensions of soundscapes. *Acta Acustica*, 6, 46.