MADRID
inter.noise 2019
June 16 - 19
NOISE CONTROL FOR A BETTER ENVIRONMENT

Investigation on Ecological Validity within Higher Order Ambisonics Reproductions of Wind Turbine Noisescapes

Bergner, Jakob[1]
Leibniz University Hannover - Institute of Communication Technology
Appelstr. 9A, 30167 Hannover, Germany

Preihs, Stephan
Leibniz University Hannover - Institute of Communication Technology
Appelstr. 9A, 30167 Hannover, Germany

Peissig, Jürgen
Leibniz University Hannover - Institute of Communication Technology
Appelstr. 9A, 30167 Hannover, Germany

ABSTRACT

Spatial sound field recordings and subsequent reproduction became more popular recently due to the availability of high quality microphone arrays, improved computational abilities and sophisticated spatial audio reproduction technologies. In the context of soundscape evaluation, valid reproducibility of the captured sound fields plays an important role. One key condition for this is, that all relevant qualitative characteristics of the soundscape are not affected by the reproduction tool chain and therefore ecological validity is preserved. This work investigates how acoustic properties might be impaired by higher-order Ambisonics soundscape reproduction. Therefore, an artificial noisescape was recorded with a spherical microphone array, coded into 4th-order Ambisonics and reproduced with a dedicated loudspeaker system. Both the original and the reproduced noisescape were subsequently investigated in terms of acoustic properties that have an influence on the psychoacoustic perception. This investigation contributes to the interdisciplinary research project WEA-Akzeptanz which aims at a holistic characterization of sound emission, propagation and perception of wind turbine noise.

Keywords: Soundscape Reproduction, Ambisonics, Wind Turbine Noise, Psychoacoustic Evaluation
I-INCE Classification of Subject Number: 66, 73, 76, 79
(see http://i-ince.org/files/data/classification.pdf)

[1]jakob.bergner@ikt.uni-hannover.de

1. INTRODUCTION

Soundscapes and their noise related equivalents, the noisescapes, are auditory scenes comprising distinct sound sources, ambient or environmental sound as well as room acoustical influences whereupon the overall auditory impression is made up by the superposition of all components. Therefore, the resulting human perception also implies various dimensions such as timbre, sound source localization and spatial envelopment which again affects human cognition and emotion with both pleasant and unpleasant aspects. The development of a model for the perceptual annoyance caused by wind turbine noise is thus one of the main goals within the interdisciplinary research project *WEA-Akzeptanz*. This task implies large scale perceptual studies under laboratory conditions of reproducible noisescapes, which are based on real world recordings, synthesized noisescapes and compositions of both. The results of those tests will be used for the training and validation of the annoyance model. To comply with the quality criteria for empirical studies, namely objectivity, reliability and validity, it is crucial to design the test conditions with care. This work focuses on the external or ecological validity to ensure that the presented noisescape in laboratory conditions meet the properties that would arise in comparable real world scenarios. To approach the goal of storing, manipulating, reproducing and comparing soundscapes, the necessity of a description model arises as well as a toolchain for reliable comparison. This work proposes a methodology for describing a soundscape or noisescape by means of physical and perceptive measures that aim to cover above mentioned dimensionality.

2. SOUNDFIELD RECORDING AND SYNTHESIS WITH HIGHER ORDER AMBISONICS

For the synthetic creation of such spatial auditory scene different technologies and approaches were developed in the past. Formerly driven by the entertainment and cinema industry, multichannel surround sound paradigms became popular and state of the art. Apart from channel-based audio formats that are motivated by psychoacoustic phenomena like phantom sound sources, two technologies were developed that aim toward a physically motivated approach, namely *Wave Field Synthesis* and *Higher Order Ambisonics* (hereinafter called Ambisonics). Ambisonics is based on the assumption that a source-free sound field can be represented as the superposition of weighted spatial basis functions, the so called spherical harmonics. The theoretical approach relies on a solution of the three dimensional wave equation which can be expressed as

$$P(\mathbf{x}, \omega) = \sum_{n=1}^{\infty} \sum_{m=-n}^{n} p_{nm}(\omega)\, b_n(r, \omega)\, Y_{nm}(\theta, \varphi) \tag{1}$$

where P is the sound pressure at position $\mathbf{x} = [x, y, z]^T$, $b_n(r, \omega)$ the radial basis function of order n, $Y_{nm}(\theta, \varphi)$ the spherical harmonics, and $p_{nm}(\omega)$ the spherical harmonic coefficients in dependency of the frequency ω. The inverse operation of Equation 1 is the decomposition of the sound field into its spherical harmonic domain described by

$$p_{nm}(\omega) = \frac{1}{b_n(r, \omega)} \cdot \int_0^{2\pi} \int_0^{\pi} P(\mathbf{x}, \omega)\, Y_{nm}(\theta, \varphi)^* \sin\theta \mathrm{d}\theta \mathrm{d}\varphi \tag{2}$$

where $\int_0^{2\pi} \int_0^{\pi} \sin\theta \mathrm{d}\theta \mathrm{d}\varphi$ denotes integration over a unit sphere and * expresses the complex conjugate. In practical realizations of this approach the sound field has to be captured and reproduced by discrete microphone and loudspeaker arrays, thus the integral in Equation 2 becomes a discrete summation and the infinite sum in Equation 1 becomes order-limited. These practical limitations introduce truncation and spatial aliasing artifacts that arise above a certain frequency. The correspondent relationships are discussed manifold for example in [1] [2]. Since the computational abilities have reached a sufficient level for the necessary complex multichannel audio signal processing in real time, Ambisonics gets more popular in various fields of research and, mostly as first order Ambisonics, even in consumer products such as virtual reality applications. These days it is possible to find high quality commercial microphone arrays as well as commercial and open source software frameworks to record and reproduce acoustic scenes in first and higher order Ambisonics. Even though this technology has reached a sufficient quality level to be application-ready, research in acoustics and psychoacoustics is still necessary to reduce, and potentially overcome the physical artifacts mentioned before. For the reproduction and subsequent evaluation of soundscapes, the requirements on the capture, processing and reproduction tool chain are very high since artifacts may alter the auditory perception notably. One approach to overcome this task is to enrich the spherical harmonic representation by additional psychoacoustic, binaural and spectral cues and thus retaining the main perceptive impression.

3. RELEVANT PSYCHOACOUSTIC MEASURES

Since the interaction and interdependence of self-contained properties within the auditory system make up the overall perception, it is important to include a reasonable selection of dimensions and cues for a descriptive model. Psychoacoustic research has identified several single properties each covering a certain aspect of auditory perception that can be derived from measurable signal-based calculations. However it is important to note that each cue is a result from both physical and empirical considerations and aims towards an approximation of a single auditory sensation or perception. Thus, when describing an acoustic scene by means of a set of signal-derived measures, the overall auditory perception can only be approximated. Keeping that in mind, we define a set of cues that cover spectral, temporal, binaural and spatial dimensions of auditory perception. The reproduction of soundscapes for subjective evaluation aims towards a plausible impression rather than physical correctness. Thus, the selection of relevant signal-based psychoacoustic measures has to be put on higher level of abstraction to get an overall view and provide meaningful comparability. For this reason low level cues that may depend on both time and frequency are condensed carefully and are thus elevated to higher level features. In the end a high level aggregation of features is aspired that serves as descriptive model of an acoustic scene as it will be elaborated in the following sections.

3.3.1. Tempo-spectral cues

The gammatone filter bank [3] is an established method for representing the frequency selective behavior of human auditory perception. The center frequencies are linearly spaced on the equivalent rectangular bandwidth (ERB) scale and bandwidth can be specified according to the desired spectral resolution. For non-static acoustic scenes, i.e. where sound sources either move or emit non-stationary signals, time-dependent analysis must be applied. For that the time-integrating behavior of auditory perception can be exploited, stating that perception relies on short frames of signals rather than on continuous processing. Each analysis frame t consist of a short range of samples, usually in the range from 10 to 100 ms, and may again be devided by a number of short time Fourier transform (STFT) windows if necessary. As suggested in [4] this discretization in time and frequency can be aggregated into a *rate map* which represents the auditory nerve firing rate and serves well as auditory spectrogram. The rate map as underlying low level feature serves then as basis for more condensed cues. Regarding the frequency dependence, two high level features are selected for this analysis, the *spectral centroid* and *spread* [5]. The centroid represents the timbre of a signal and is defined as centre of gravity of the current frame's magnitude spectrum $X_t(k)$

$$c(t) = \frac{\sum_{k=1}^{K_N} |X_t(k)| \cdot f_k}{\sum_{k=1}^{K_N} |X_t(k)|} \tag{3}$$

where t denotes the current temporal frame and summation is applied over the discrete frequencies f_k up to the nyquist frequency at bin $k = K_N$. Spectral spread on the other hand represents the deviation from the centroid

$$s(t) = \sqrt{\frac{\sum_{k=1}^{K_N} (f_k - c_t)^2 \, |X_t(k)|}{\sum_{k=1}^{K_N} |X_t(k)|}} \quad , \tag{4}$$

which is small for narrowband and high for broadband signals. Further information on tonality properties are provided by the spectral flatness measure SFM [6] which is calculated by the ratio of geometrical and arithmetical mean of the spectrum

$$\mathrm{SFM}(t) = \frac{\left(\prod_{k=1}^{K_N} |X_t(k)|\right)^{1/K_N}}{\frac{1}{K_N} \sum_{k=1}^{K_N} |X_t(k)|} \tag{5}$$

and provides values between 0 (peaky, tonal spectrum) and 1 (noise-like spectrum).

3.3.2. Binaural cues

Research of spatial hearing [7] identified the interaural time and level differences ITD and ILD as important cues for localization of sound sources. Thus, for the description of an acoustic scene exhibiting distinct sound sources, these cues deliver relevant information. Both measures are derived from the interaural cross correlation function *ICC* and are calculated for each frequency band and time frame. The implementation used here follows the suggestions in [8] that differ from previous definitions towards sound source localization applications. It has to be noted, that the superposition of separate sound sources with coherent frequency parts may influence the ILD and ITD in a interfering way leading to reduced values compared to isolated investigations of each sound source individually. In the field of acoustic scene analysis it can be stated that strong ILD and ITD can be interpreted as the presence of a dominant sound source.

Analogous to the tempo-spectral cues, the feature condensing along the frequency dimension is performed by amplitude weighting within the current temporal frame

$$\mathrm{ITD}\,(t) = \frac{\sum_{k=1}^{K_N} |X_t(k)| \cdot ITD_t(k)}{\sum_{k=1}^{K_N} |X_t(k)|} \tag{6}$$

$$\mathrm{ILD}\,(t) = \frac{\sum_{k=1}^{K_N} |X_t(k)| \cdot ILD_t(k)}{\sum_{k=1}^{K_N} |X_t(k)|} \tag{7}$$

3.3.3. Spatial cues

Another very important dimension of the perception of acoustic scenes is the general spatial and envelopment impression. As mentioned before, a natural acoustic scene is composed by distinct, locatable sound sources, ambient sound and background noise sources without clear localization as well as a room acoustic environment with a certain reflection and envelopment pattern. The reconstruction of latter is an important task when it comes to creating plausible, synthetic acoustic scenes. Typical room acoustic measures such as the reverberation time T_{60} or the direct-to-reverberant ratio DRR provide valuable information for the spatial perception of acoustic scenes. However the in-situ deduction of these measures is not straightforward because the direct sound estimation is ambiguous without knowledge of type and location of the sound source(s). Nevertheless the estimations of the DRR is feasible by means of directional analysis with microphone arrays for which different approaches and theoretical considerations exist. In general, the sound from the dominant direction of arrival (DOA) is expected the direct sound and its energy is subsequently put in relation to the overall sound energy. One method to estimate the DOA is based on the directive sound intensity at the receiver position. The intensity vector $\mathbf{I} = [I_x, I_y, I_Z]^T = p \cdot \mathbf{u}$ with sound pressure p and velocity vector $\mathbf{u} = [u_x, u_y u_z]^T$ can be retrieved by first order Ambisonics signals commonly referred to as B-format as described in [9]. In this case and according to Equation 2 the following approximations can be applied:

$$\begin{aligned}
p &= \sqrt{2} p_{nm} && \text{with } n = 0,\ m = 0 \\
u_x &= p_{nm} && \text{with } n = 1,\ m = 1 \\
u_y &= p_{nm} && \text{with } n = 1,\ m = -1 \\
u_z &= p_{nm} && \text{with } n = 1,\ m = 0
\end{aligned} \tag{8}$$

This implies the use of a microphone array, that is suitable for Ambisonics decomposition such as a tetrahedral microphone or multiple sensors mounted into a rigid sphere. The direction of arrival of the the dominant sound is then the opposite of the intensity vector direction while the length of the intensity represents the dominance of the sound source. Investigations have shown that this approach is subject to strong influences of artifacts which is why a robust detection of the direction-of-arrival is not feasible. Another straightforward approach utilizes an array of multiple microphones that show some sort of directional behavior, either by its conversion principle like in cardioid microphones or

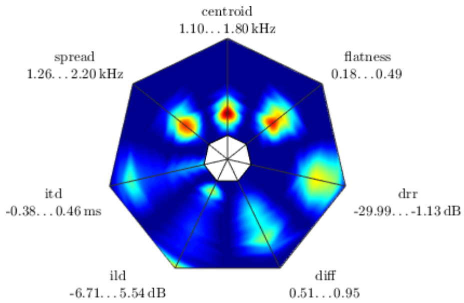


Figure 1: Exemplary soundscape summary diagram.

by mounting them onto the surface of a rigid sphere. With a close to equal distribution of the microphones on the sphere the direction of arrival can be estimated by weighting the microphone look directions with their respective signal magnitude as suggested in [10]

$$\mathbf{r}_{DOA}(k, t_f) = \sum_{q=1}^{Q} |x_q(k, t_f)| \cdot \mathbf{n}_q \tag{9}$$

where $r = [r_x, r_y, r_z]^T$ is the vector pointing towards the direction of arrival within the current STFT window t_f, Q the number of microphones and $\mathbf{n}_q = [n_{x,q}, n_{y,q}, n_{z,q}]^T$ the unit vector pointing from origin to the qth microphone of the array. Even though the magnitude of the resulting vector must not be mistaken as sound intensity $||\mathbf{r}_{DOA}(k,t)|| \neq || - \mathbf{I}(k,t)||$ the approximation of the direct sound energy portion can be conducted with its help as suggested by [10] under the proposition that $r_{DOA}(k,t)$ inhibits temporal averaging over several STFT windows t_f within a time frame t

$$\mathbf{r}_{DOA}(k,t) = \mathbf{E}\left[\frac{\mathbf{r}_{DOA}(k,t_f)}{||\mathbf{r}_{DOA}(k,t_f)||}\right] \tag{10}$$

for which the direction of arrival can be assumed to be static. In this case diffuseness can be expressed as

$$\psi(k,t) = 1 - ||\mathbf{r}_{DOA}(k,t)|| \tag{11}$$

and direct to reverberation ratio DRR to

$$DRR(k,t) = \frac{1}{\psi} - 1 = \frac{||\mathbf{r}_{DOA}(k,t)||}{1 - ||\mathbf{r}_{DOA}(k,t)||} \quad . \tag{12}$$

3.3.4. Soundscape summary diagram

Different cues were identified in the previous section that cover selected properties regarding the acoustic dimensions timbre, (binaural) direction of arrival as well as diffuseness. We propose that these dimensions hold a selection of relevant information that are needed for the description of acoustic scenes. Thus, being able to reconstruct these features will bring us to the goal of plausible soundscapes much closer. For comparison reasons and for intuitive evaluation we suggest to aggregate the detailed cues into a more condensed representation by replacing the time variance with its discrete temporal distribution. In this way, histograms of the individual characteristics are obtained that can be put together for a general overview. Figure 1 shows the distinct histograms of the correspondent cues along the axes with relative occurence coded in colors. The numbers below the axes labels denote the complete dynamic range of the shown respective cue, starting from the minimum value just outside the white center up the maximum value at the outer end.

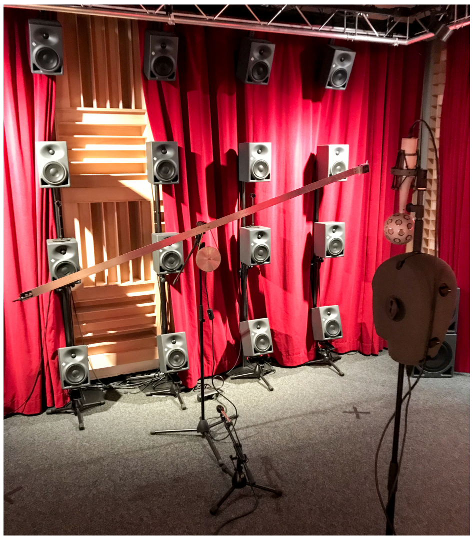
4. EXPERIMENTAL LAYOUT AND DATA BASIS

The previously proposed analysis of soundscapes was applied to a range of controllable test conditions in the course of an experiment. For that different reproducible soundscapes were generated, recorded and resynthesized. The comparison of the original and reproduced soundscape by means of the proposed analysis chain is then intended to provide indications for ecological validity. To approximate a wind turbine scenario, the generated noisescape consists of a circularly moving sound source, which was realized by a miniature loudspeaker (Mini MusicMan Soundstation) playing a white noise signal on a custom tailored turning device with a diameter of two meters which rotates at a turning rate of approximately $12.4\,\mathrm{min}^{-1}$ as shown in Figure 2a. A reference microphone was used to exactly determine the position of the loudspeaker during the measurement. For the capturing of the noisescape a binaural dummyhead (Neumann KU100) and a spherical microphone array (mhacoustics Eigenmike) were utilized simultaneously. In order to maintain almost the same directional properties of the two microphones, the Eigenmike was mounted upside down directly on top of the dummyhead. This setup was a tradeoff, since the close proximity between the microphones introduces unwanted reflections especially on the lower Eigenmike sensors. The dummyhead was aligned such that the the rotating center of the turning device corresponds to the ear level, while the distance was 1.6 m chosen to have a similar angular projection than a real measurement of a wind turbine in 100 m distance. All signal were recorded synchronously at a sample rate of 48 kHz with 24 bits per sample. For the analysis specimen of 20 s were extracted each with the same starting point where the rotating device points downwards. The Ambisonics resynthesis was conducted in the *Immersive Media Lab* [11] at the Institute of Communication Technology. For the Ambisonics encoding of the Eigenmike signals the mhacoustics VST-plugins were utilized while for the decoding and rendering the Ambisonics plugin suite by the Insitute of Electronic Music and Acoustics IEM was used which employs the all-round Ambisonics decoder [12]. The loudspeaker layout consisted of 36 spatially distributed Neumann KH120 Loudspeakers where the front pane is densly equipped with 5x4 loudspeakers as can be seen in Figure 2b). The resynthesis was then recorded with the same microphone setup for comparability. To compare two different room acoustic conditions, the noisescape was established and recorded under anechoic conditions (Figure 2a) as well as under reverberant conditions ($T_{30} \approx 0.3$ s) (Figure 2b).

5. RESULTS

In order to compare an original soundscape with its Ambisonics resynthesis the measures defined in the previous sections are applied to the recordings. Since physical artifacts are expected, the main focus of the comparison is to find similarities and differences to identify properties of the Ambisonics resynthesis that have to be improved. The analysis of the results is in the first instance applied to single features that represent the dimensions timbre, binaural localization, direction of arrival and diffuseness. Figure 3 shows the spectral properties of the original and resynthesized soundscapes. The left column contains the analysis for the soundscape recorded under anechoic conditions and the right column represents the same soundscape in reverberant conditions. Each column shows the analysis of both original and resynthesized recording. By looking at the original conditions (blue and yellow curves) it can be seen that the centroid, spread and flatness show only little variance and no periodic behaviour at all. This is the

a



b

Figure 2: Laboratory setups for recordings of a moving sound source. a) Anechoic recording. b) Recording of a moving sound source and setup for loudspeaker resynthesis.

expected behaviour for a stationary noise-like signal with no or little room acoustical influence. At the same time, the resynthesized soundscapes show more fluctuations over time and even some periodic similarities in both conditions for almost all presented features. Since the periodicity matches the rotation of the sound source it can be said that the resynthesis of the recorded sound source is altered spectrally for distinct positions. The binaural cues used for localization are shown in Figure 4 in the same way. The interaural time difference ITD can be reconstructed quite well as the top row shows. Even though the original soundscape produces clearer ITDs also within a larger range, the resynthesis shows obvious similarity. The interaural level difference however in the bottom row shows much less distinct results for the resynthesis. Considering that the ILD is more effective in higher frequencies, this behaviour can partly explained with spatial aliasing above the theoretical aliasing frequency of the Eigenmike of $f_{alias} \approx 5200\,\mathrm{Hz}$. Another explanation for low level differences between the left and right hemisphere is a higher amount of diffuse sound. This aspect will be discussed below. Beside the binaural cues for sound source localization another approach to detect the direction of the dominant incoming sound is shown in Figure 5. It is derived from the spherical array processing discussed in section 3.3 and provides information based on the sound intensity vector at the receiver position. It can be seen that the horizontal direction of arrival (top row) can be reproduced quite well for the resynthesized soundscapes. The elevation detection on the other side is more ambiguous as can be seen when comparing the resynthesis of the anechoic soundscape with the reverberant one. Latter shows good similarity in tendency with the original soundscape while the resynthesis of the anechoic soundscape shows almost no elevational variance at all. This unexpected behaviour has to be investigated further, as it appears to be a systematic failure. The last dimension, the diffuseness, shows high deviations in the current system for soundscape reproduction and has therefore high priority for improvements for future work. Figure 6 shows the analysis results for the diffuseness and the direct-to-reverberation ratio as described in section 3.3. For this dimension the presented experiment layout is beneficial since two

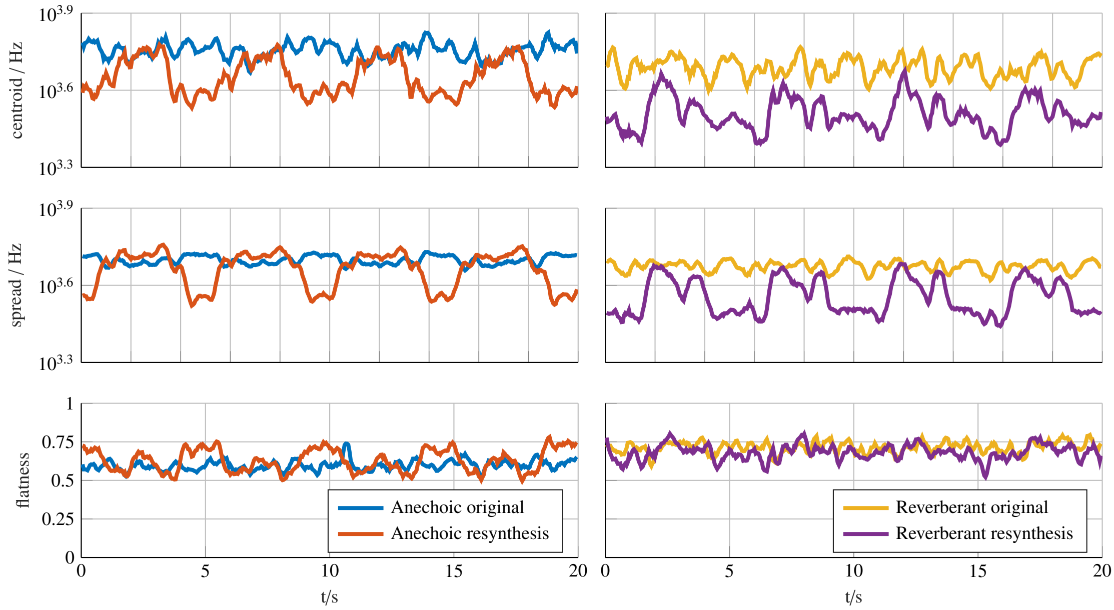


Figure 3: Spectral analysis by means of centroid, spread and flatness (top to bottom).

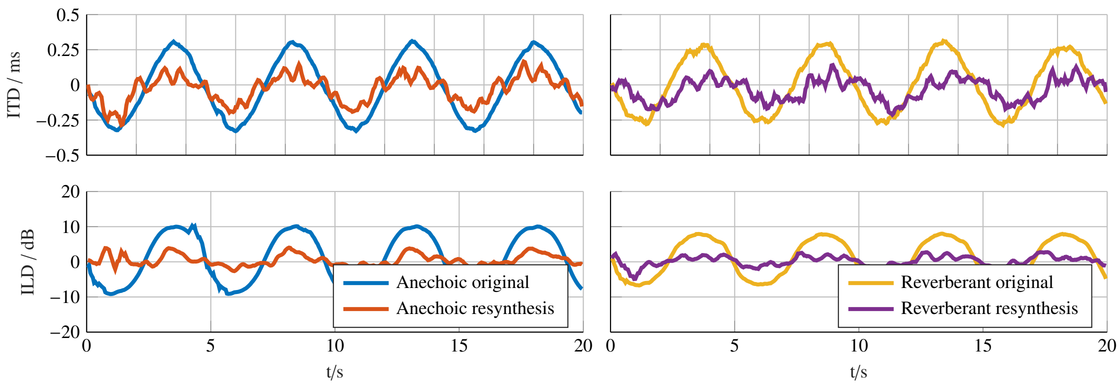


Figure 4: Analysis of the binaural cues ITD (top) and ILD (bottom).

very different room acoustical conditions are used to evaluate the methodology. As expected, the two original soundscapes in the anechoic and reverberant environment (blue and yellow respectively) differ in the general level of these features. Apparently the original anechoic soundscape shows slight periodic variation over time which means that diffuseness varies with sound source location. This can be expected to result from the microphone placement on top of each other which introduces reflections from the dummy head to the spherical microphone array. At the same time the results show the weakness of the exploited Ambisonics resynthesis in reproducing distinct, non reverberant sound sources. The diffuseness level of the resynthesis of the anechoic soundscape is comparable with the level of the reverberant soundscape resynthesis (red and purple curve respectively in the top row) even though the original levels are quite different. The same behaviour can be observed for the DRR in the bottom row. It can be concluded that the current Ambisonics reproduction tool chain introduces additional reverberation and diffuseness regardless of the diffuseness of the recorded soundscape. To be able to commit a statement of the auditory properties of a soundscape, the previous shown cues are now aggregated into a time independent representation. Its purpose is to compare the various dimensions of a soundscape in a comprehensive way and to identify

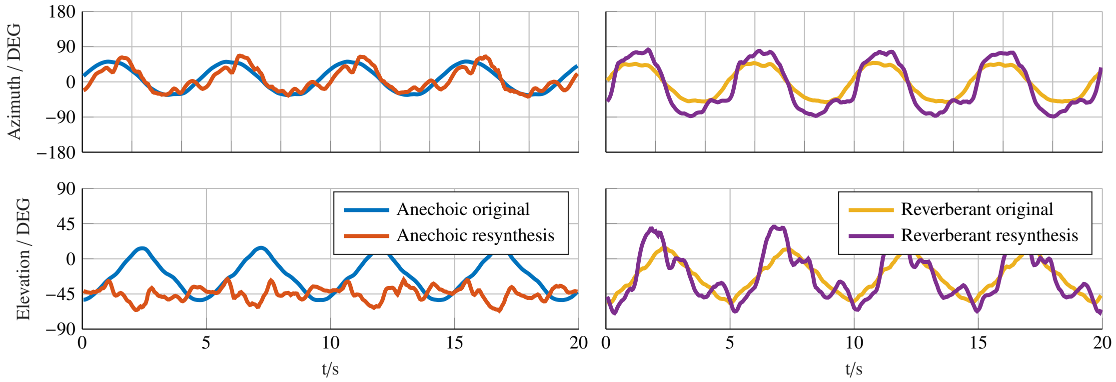


Figure 5: Analysis of the direction of arrival in horizontal (top) and vertical (bottom) direction derived by spherical microphone array processing.

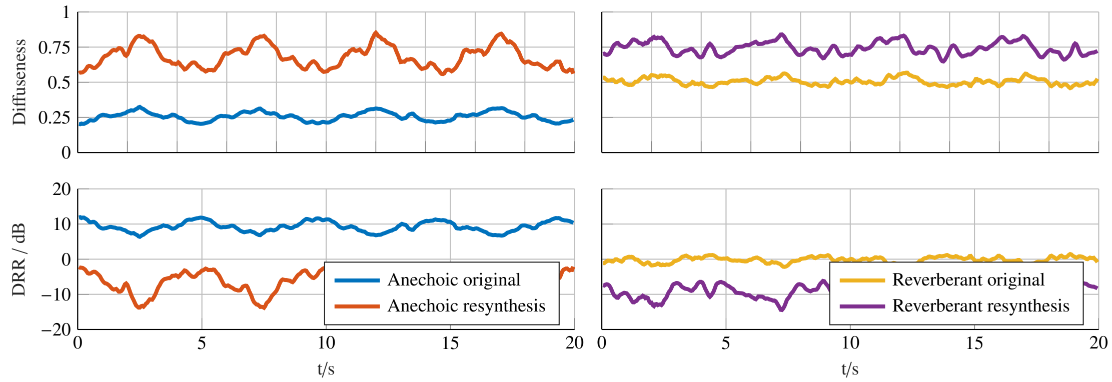


Figure 6: Analysis of the spatial dimension by means of diffuseness (top) and direct to reverberation ration (bottom).

similarities and differences. Figure 7 shows thus the summary of the four investigated soundscape conditions. Some general finding can be obtained by that.

– The spectral cues centroid, spread and flatness of the resynthesized soundscapes show more variance and thus are not as stable as the the original soundscapes and as would be expected for stationary, noise-like signals.

– The binaural cues ITD and ILD show less variance within the resynthesized soundscape which can be interpreted as less ditinct binaural localization.

– The spherical cues diffuseness and DRR show similar distributions, however the resynthesized soundscapes are located on a much more reverberant level.

This categorized and condensed results provide information on how to improve the Ambisonics soundscape reproduction in future.

6. CONCLUSIONS

This work presents investigations on how properties of human perception might be impaired by the use of higher order Ambisonics technologies for soundscape reproduction. For this purpose a methodology is proposed that aims to describe the general perceptive impression of a soundscape and which is based on individual low level

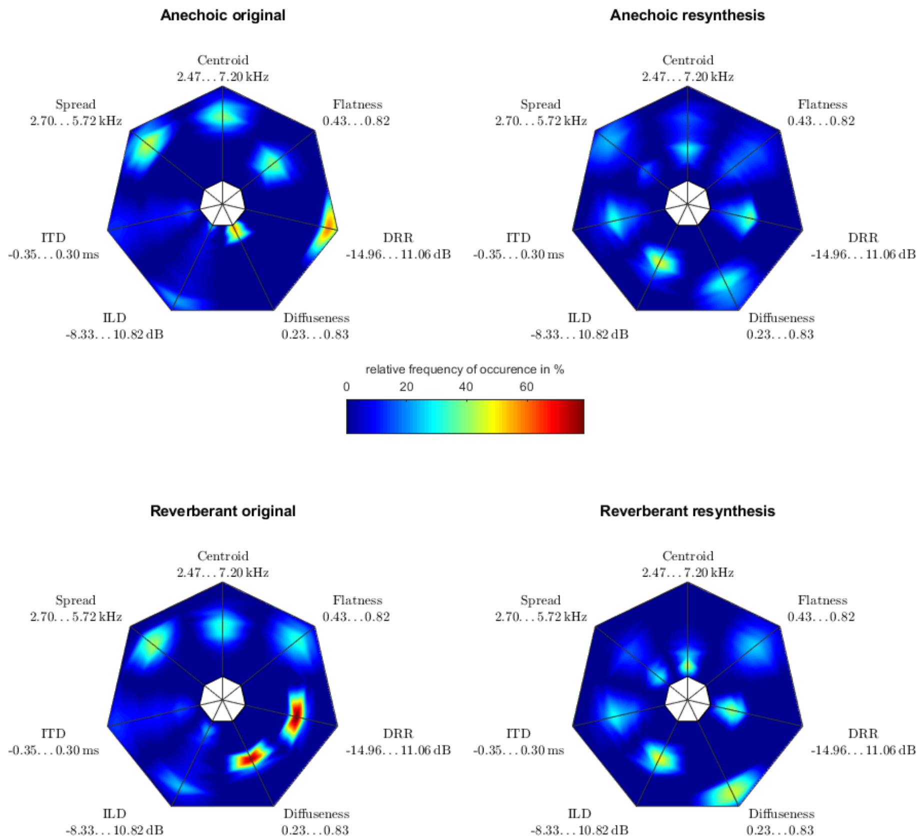


Figure 7: Aggregated cues for soundscape description.

cues that describe certain aspects of the various dimensions of auditory perception. To apply and evaluate this methodology an experiment was laid out in which a reproducible noisescape is recorded with a spherical microphone array and reproduced by means of Ambisonics processing. The comparison of the original and the resynthesized noisescape was conducted in terms of the presented methodology. Thus, this experiment serves as testbed for the methodology itself as well as a measurement tool for the current soundscape reproduction infrastructure. The results show that the current Ambisonics reproduction exhibit notably overall alterations in the observed dimensions timbre, sound source localization and diffuseness. At the same time an in-dept investigation of the single cues show in detail similarities especially for the binaural cues ILD and ITD and the direction of arrival. One suggestion for future improvements is to use these cues to separate the dominant sound source and emphasize it during reproduction by means of adequate signal processing. It can be expected that by this action the localization and the diffuseness can be improved towards the propertties of the original noisescape. The spectral deviations represented by centroid, spread and flatness may be adapted by meaningful equalization of the reproduced signal and by detecting and reducing possible room acoustic causes of errors. In general, the proposed methodology of a soundscape description provide valuable and detailed information on the technical abilities of Ambisonics soundscape reproduction.

7. ACKNOWLEDGEMENTS

The project WEA-Akzeptanz is funded by the German Federal Ministry for Economic Affairs and Energy under grant number 0324134A.

8. REFERENCES

[1] Boaz Rafaely. Analysis and design of spherical microphone arrays. *IEEE Transactions on Speech and Audio Processing*, 13(1):135–143, jan 2005.

[2] Boaz Rafaely, Barak Weiss, and Eitan Bachmat. Spatial Aliasing in Spherical Microphone Arrays. *IEEE Transactions on Signal Processing*, 55(3):1003–1010, mar 2007.

[3] Brian R. Glasberg and Brian C.J. Moore. Derivation of auditory filter shapes from notched-noise data. *Heraring Reseach*, 47(1):103–138, 1990.

[4] Guy J. Brown and Martin Cooke. Computational auditory scene analysis. *Computer Speech and Language*, 8(4):297–336, 1994.

[5] George Tzanetakis and Perry Cook. Musical Genre Classification of Audio Signals. *IEEE Transactions on Speech and Audio Processing*, 10(5):293–302, 2002.

[6] Geoffroy Peeters, Bruno L. Giordano, Patrick Susini, Nicolas Misdariis, and Stephen McAdams. The Timbre Toolbox: Extracting audio descriptors from musical signals. *The Journal of the Acoustical Society of America*, 130(5):2902–2916, 2011.

[7] Jens Blauert. *Spatial Hearing*. The MIT Press, Harvard, MA, revised ed edition, 1997.

[8] Jonas Braasch. A precedence effect model to simulate localization dominance using an adaptive, stimulus parameter-based inhibition process. *The Journal of the Acoustical Society of America*, 134(1):420–435, 2013.

[9] Ville Pulkki. Applications of directional audio coding in audio. In *Proc. of the 19th ICA*, pages 1–6, 2007.

[10] Archontis Politis, Symeon Delikaris-Manias, and Ville Pulkki. Direction-of-Arrival and Diffuseness Estimation Above Spatial Aliasing for Symmetrical Directional Microphone Arrays. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6–10. IEEE, 2015.

[11] Robert Hupke, James Ordner, Jakob Bergner, Marcel Nophut, Stephan Preihs, and Juergen Peissig. Towards a Virtual Audiovisual Environment for Interactive 3D Audio Productions. In *AES International Conference on Immersive and Interactive Audio*, 2019.

[12] Franz Zotter, Matthias Frank, and A E S Student Member. All-Round Ambisonic Panning and Decoding. *Journal of the Audio Engineering Society*, 60(10):807–820, 2012.

[13] Eric Scheirer and Malcolm Slaney. Construction and evaluation of a robust multifeature speech/music discriminator. In *Proc. of the IEEE ICASSP*, pages 1331–1334, 1997.

[14] Ville Pulkki. Spatial sound reproduction with directional audio coding. *Journal of the Audio Engineering Society*, 55(6):503–516, 2007.